5-2016

# Three Essays on Big-Box Retailers and Regional Economics

Denis Peralta

*Utah State University*

THREE ESSAYS ON BIG-BOX RETAILERS AND REGIONAL ECONOMICS

by

Denis Peralta

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Economics

Approved:

| | |
|---|---|
| Dr. Man-Keun Kim<br>Major Professor | Dr. DeeVon Bailey<br>Committee Member |
| Dr. Reza Oladi<br>Committee Member | Dr. Ryan Bosworth<br>Committee Member |
| Dr. Vijay Kannan<br>Committee Member | Dr. Mark McLellan<br>Vice President for Research and<br>Dean of the School of Graduate Studies |

UTAH STATE UNIVERSITY
Logan, Utah

2016

# ABSTRACT

Three Essays on Big-Box Retailers and Regional Economics

by

Denis Peralta, Doctor of Philosophy

Utah State University, 2016

Major Professor: Man-Keun Kim
Department: Applied Economics

The big-box retail stores such as Wal-Mart and Target have become the focus of many studies researching their impacts on local economic outcomes. This dissertation studies three related topics: (i) the dynamic interrelationship among the presence of the big-box stores, retail wage, and employment, (ii) the impact of the big-box retailers on personal income growth, and (iii) the dynamic interrelationship between the presence of big-box retailers and personal income growth. The research draws important insights with potential implications for regional developers and policy makers.

The first essay analyzes the dynamic relationship among the presence of the big-box retailers, retail wage, and employment at the county level for 1986-2005. A vector autoregression model is applied on panel data. Impulse response functions and variance decompositions are also presented. Results suggest that the presence of big-box stores decreases retail wages and increases retail employment. Retail employment has a higher impact on the retailers' location decision than retail wage. The results also show that the presence of Wal-Mart drives the above-mentioned effects, while the presence of Target is insignificant.

The second essay investigates the impact from the presence of big-box retailers on personal income growth in U.S. counties between 2000 and 2005 - based on neoclassical growth models of cross-country income convergence. Results suggest that counties having both Wal-Mart and Target stores experienced slower growth in personal income. After controlling for spatial autocorrelation, similar to the first essay, the effect of Wal-Mart's presence on personal income growth is dominant in terms of statistical significance relative to Target's.

The third essay expands the second essay and investigates the dynamic interaction between the presence of big-box retailers and personal income growth over time at the county level for the period 1987-2005, using a panel vector autoregression model. For this analysis, the earning shares of natural resources and manufacturing sectors are included - assuming that all the variables are endogenous to one another. The findings indicate that big-box retailers negatively affect personal income growth, which is consistent with the second essay. However, personal income growth has an insignificant effect on the big-box retailers' location decision.

(100 pages)

PUBLIC ABSTRACT

Three Essays on Big-Box Retailers and Regional Economics

Denis Peralta

Throughout the years, big-box retail stores such as Wal-Mart and Target have become the focus of many studies researching their impacts on local economic outcomes (i.e. employment, wages, poverty level, food prices, etc.) within specific regions, states, counties and localities in the U.S. This dissertation covers three closely related topics in regional science: (i) the dynamic interrelationship among the presence of the big-box stores, retail wage, and employment, (ii) the impact of the big-box retailers on personal income growth, and (iii) the dynamic interrelationship between the presence of the big-box retailers and personal income growth. The research draws important insights with potential implications for regional developers and policy makers.

The work builds on previous literature and advanced statistical approaches such as the panel vector autoregression (panel – VAR) model and spatial econometrics. The empirical results suggest that: (i) the presence of big-box retailers increases retail jobs while it decreases retail wages. Wal-Mart seems to drive the effects while Target's presence appears inconsequential. (ii) Counties with big-box retailers experienced slower growth in personal income between 2000 and 2005. After controlling for spatial dependence, the impact of Wal-Mart's presence remains negative and significant while Target's effect becomes insignificant, and (iii), big-box retailers have a negative impact on personal income growth over time, whereas personal income growth has an inconsequential effect in the number of big-box retailers in the region.

# ACKNOWLEDGMENTS

I thank my creator for giving me the strength and determination to pursue and complete this dissertation. I also express my genuine gratitude and appreciation to my advisor, Dr. Man-Keun Kim, for all his guidance, and mentoring. Your patience, support and encouragement made this dissertation a reality.

I am grateful to the members of my committee; Dr. DeeVon Baily, Dr. Reza Oladi, Dr. Ryan Bosworth, Dr. Vijay Kannan and also in the APEC Department; Dr. Hernan Tejeda. All of your helpful suggestions, comments and constructive criticism made this work possible.

I am indebted to my family, friends, and classmates for their encouragement, moral support and patience as I worked through the entire research process to this final document. Special thanks goes to my wife Bernadette Peralta for your love, support, motivation and encouragement throughout this long journey.

Denis J. Peralta

CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

During the last two decades, big-box retail stores such as Wal-Mart and Target have become the focus of a series of studies researching their impacts on local economic outcomes (i.e. employment, wages, poverty level, food prices, etc.) within specific regions, states, counties and localities in the U.S. The sizable growth and expansion of these big-box retailers, especially Wal-Mart, have drawn significant attention from the media, other retailers, local policymakers and academics (Bonanno and Goetz, 2012). Numerous studies listed in chapter 2, the literature review section, suggest that big-box retailers have shown positive, negative and mixed effects on local economic outcomes (i.e. employment, wages, prices, poverty, etc.). Empirical results are heavily dependent on data used, methodology adopted, regional specification, and scope of the study. The goal of the essays in this dissertation is to examine the economic impact of these big-box retailers on the local economies of U.S. counties by overcoming prevailing endogeneity problems in the previous literature (first essay), addressing research gaps in the literature, and studying the relationship between big-box retailers and income growth (second and third essays, respectively). Findings from this research can provide beneficial insights and may have important implications for local residents, policy makers and researchers.

The first essay analyzes the impact of big-box retailers Wal-Mart and Target, on retail employment and retail wage at the county level within the 48 contiguous states for the period 1986-2005. The dynamic interrelationship among the variables of interest is examined utilizing the vector autoregression model on panel data (panel VAR). The panel VAR allows for a county-specific unobserved heterogeneity in the variables, i.e., fixed effects. In addition, the panel VAR model does not require strong assumptions that are necessary in other models that may use questionable instruments to control for endogeneity among variables, but rather assumes all variables in the system to be endogenous to each other. Moreover, the model permits the calculation of Impulse response functions (IRF), which allow separating the dynamic responses of retail employment, wage, and the big-box stores to particular shocks from each of these variables of interest. Additionally, variance decompositions are analyzed, which give the total variation contribution on a particular variable as a result of a shock on another variable.

Empirical results show that the presence of big-box retailers increases the number of retail jobs while it decreases the level of retail wages. The total effect on retail employment is relatively larger than the effect on retail wages. These effects are mainly driven by the presence of Wal-Mart stores rather than by Target stores. The impact of Target stores on retail wage and employment seem inconsequential, as they are rather small and statistically insignificant. Target's location decision seems to be slightly more heavily affected by retail wage than employment level. On the other hand, retail employment has a greater impact on Wal-Mart's location decision. As in Basker (2007), Target is portrayed as having a follower strategy when it comes to location decision in counties that already have a Wal-Mart store. On the contrary, Wal-Mart may choose to avoid allocating in counties that already have a Target store.

The second essay investigates the impact of these big-box retailers on personal income growth in counties, using the cross-sectional county level data between 2000 and 2005 for the 48 contiguous states. The study is built upon neoclassical growth models of cross-country income convergence. Various model specifications are estimated, including spatial models that control for spatial dependences in the analysis. The objective of the second essay is to determine if there is a relationship between personal income growth and the presence of big-box retail stores. The results indicate that counties that have both Wal-Mart and Target stores have experienced slower growth in personal income. After controlling for spatial autocorrelation, similar to the first essay, the effect of Wal-Mart on personal income growth is dominant in terms of statistical significance relative to Target's.

Potential endogeneity between Wal-Mart and Target location decisions and local economic outcomes, may be a source of misspecification when examining the effect of these big-box retailers on the local economy. Bonanno and Goetz (2012) emphasize this potential endogeneity and misspecification issue. Therefore, the third essay, investigates the dynamic endogenous inter-relationship among the big-box retailers and personal income growth using county level data for the period 1987-2005. Given that natural resource endowment and the structure of the economy are important elements in income growth, the earnings share of natural resources and manufacturing sectors are likewise included - assuming that all the variables in the new system are endogenous to one another. Similar to the first essay, the panel VAR approach is utilized to examine the dynamic interaction among the variables of interest. The findings indicate that big-box retailers have a negative impact on county personal income growth, consistent with conclusions from the second essay; however, this effect quickly dissipates after the first period. On the other hand, personal income growth has an inconsequential effect on the change of the number of the big-box stores in the county.

The rest of this dissertation is organized as follows: Chapter two presents an overview of the literature regarding big-box retailers, primarily Wal-Mart, as well as the impact on the U.S. regional economy. Subsequently, an analysis of big-box retailers, local retail employment, and wages is presented in Chapter three. The study of the effect of big-box retailers on personal income growth is discussed in Chapter four. Chapter five covers the analysis of the dynamic inter-relationship among big-box retailers, personal income growth, natural resource endowment (earning shares of the natural resources sector) and the earnings share of the manufacturing sector. Chapter six provides the conclusions of this dissertation.

# CHAPTER 2

# LITERATURE ADDRESSING BIG-BOX RETAILERS AND THE REGIONAL ECONOMY

Throughout the years, big-box retail stores such as Wal-Mart and Target have become the focus of a series of studies researching the impacts of these type of firms on local economic outcomes. Their business model consists of large-footprint model buildings that combine firm-wide efficiencies, transportation and supply networks, information technology, negotiating power with suppliers, and proximity to distribution centers that enable them to sell a broad range of consumer goods at lower prices than (smaller) competing retailers (Basker, 2007, Basker, Klimek and Hoang Van, 2012, Holmes, 2011). The growth and expansion of big-box retailers' stores, especially Wal-Mart's, have drawn significant attention from the media, other retailers, local policymakers and academics (Bonanno and Goetz, 2012).

## 2.1. General Literature Review

Many studies analyze the effect of big-box retailers, usually Wal-Mart, on a series of (economic) indicators including other retail businesses, employment, wage, sales, poverty levels, prices, and local tax revenues. Most research and empirical results are mixed and heavily dependent on the data, the methodologies utilized, and regional specification. Since the aim of this research is to evaluate the effect of Wal-Mart and Target on the local economy, this section reviews previous related works and highlights a research gap in the literature.

Arguably Stone (1988) may be the first study addressing the effect of Wal-Mart's presence in towns of Iowa. The study's findings show that Wal-Mart's presence has a stronger negative effect (on the sales levels of competing businesses) in smaller towns than in larger towns.

Moreover, competitors with similar product lines as Wal-Mart stores often suffer greater losses in sales, than those offering non-competing products or services (Stone, 1988). Other studies' results agree with Stone (1988) in that the impact of Wal-Mart's entry on local retailers' sales is considered negative for direct competitors, although some complementary establishments may yield positive benefits from Wal-Mart's presence (Ailawadi, et al., 2010, Artz and Stone, 2006, Irwin and Clark, 2006, Jia, 2008, Stone, 1988, Stone, 1997, Stone, 1995).

In a related study, Ozment and Martin (1990) conclude that Wal-Mart's entry into a market location generally increases opportunities for non-competing businesses in that (market) location. In contrast, there are studies that link Wal-Mart and other large discount chains' presence in a market location to the closure of small shops in downtown and local main streets, along with declines in employment and wages, community disruption and higher poverty (Freeman, 2003, Goetz and Swaminathan, 2006, McGee and Gresham, 1996, Quinn, 2005). However, a study by Barnes, et al. (1996) which focuses on Northeast markets, concludes that neither the number of current establishments nor their sales growth are negatively affected by Wal-Mart's presence.

In a related venue, Wal-Mart's impact on local retailers may have a considerable effect on local employment and wages. Opponents of Wal-Mart stores continuously argue that its presence negatively affects local employment while depressing wages as well. Nonetheless, study results of Wal-Mart's effect on employment and wages are often conflicting. One of the first studies about Wal-Mart's effect on employment is Ketchum and Hughes (1997), which analyze 16 counties in Maine during 1990-1994. They focus on the impact of Wal-Mart's entry on per capita employment and average wage for manufacturing, retail and services sectors. They conclude that there is not a statistically significant difference in per capita employment and wages among the 16 counties with and without accounting for Wal-Mart.

Hicks and Wilburn (2001) find a modest increase in retail employment (approximately 54 workers per county) as a result of Wal-Mart presence in West Virginia. They use the county-level data for the period of 1988-2000, controlling for spatial autocorrelation in neighboring counties that likewise have Wal-Mart stores. They also find that there is no effect on retail wages. In a subsequent study, Hicks (2007) analyzes eight Pennsylvania counties that have at least one Wal-Mart store from 2001-2005. Quarterly workforce indicators are used to assess the effect of the company's entry and expansion, on the dynamics of employment and wages using a cross-sectional sample. Hicks (2007) finds that Wal-Mart's entry has no significant impact on retail wages for existing employees in the retail sector, while new hires enjoy a roughly $90 per month premium. The impact on employment is a net gain of roughly 50 jobs, which is consistent with the findings in Hicks and Wilburn (2001) and Basker (2005). Wal-Mart's expansion effect on retail wages is examined again in Hicks (2008) by looking at retail employment and aggregate employment in Maryland's 23 counties from 1988-2003. His findings show that the impact of Wal-Mart on retail employment is negative, but on retail wage it is positive. Hicks (2008) interprets these results as an increase in marginal productivity of labor.

Keil and Spector (2005) examine the effect of Wal-Mart's presence on income differentials and unemployment between blacks and whites in Alabama, using county census data for 1980 and 1990. They find that Wal-Mart's presence significantly correlates to lower unemployment for blacks, while the impact on income is trivial after controlling for other socio-economic variables. Jantzen, Pescatrice and Braunstein (2009) use cointegration techniques and causality tests to examine the relationship between Wal-Mart U.S. sales and a set of macro measures of income, employment, production, and prices. Their study is done at the national level using data with Wal-Mart sales for the periods from 2000-2005 and from 1995-2005. They conclude that Wal-Mart's sales soar during periods of slow economic growth and decline during booming periods, i.e., Wal-Mart sales move counter to overall economic conditions.

An extensive study of Wal-Mart's impact on the local economy is found in Basker (2005). The author uses a self-collected data set with Wal-Mart store locations along with county-level data for the contiguous states in the U.S. She controls for endogeneity of store location decision and local economy utilizing company-assigned store numbers as the proxy for planned store openings. While the total number of stores opened every year is treated as predetermined, the progression of the stores' numbers is used to assign a "planned year" of opening to each Wal-Mart store, for the period 1977-1999. In other words, this number is aggregated to the county level, and the number of "planned store openings" in each year, within each county (i.e., if the stores had opened in the order in which they were numbered, they would be part of the number of stores opened in a specific year) is then used as an instrument in place of the actual number of store openings. She determines that Wal-Mart's presence has an initial moderate creation of new jobs, but it dissipates over time from 100 to 50 jobs within a 5-year period; which is consistent with Hicks and Wilburn (2001). In addition to this, she finds a loss of about 30 wholesaling jobs and small growth in restaurant employment. However, as pointed out by Goetz and Swaminathan (2006), the fact that Basker (2005) does not distinguish between full and part-time employment and likewise the sample considers counties with positive employment growth and employment levels above 1,500 in 1964, may lead to sample selection bias.

Basker (2007) explores Wal-Mart's competitive (volume) advantage and how its presence affects consumer prices, local labor markets, global and local competitors, suppliers and product selection. Although the study is more of a qualitative analysis and survey of the literature on Wal-Mart, she emphasizes how Wal-Mart decision of location depends on the local economic conditions. At the same time, the study makes references on how other chain retailers including Target have changed their business practices to emulate Wal-Mart's. Notably, a significant portion of previous research has focused on examining Wal-Mart's effect on the local economy given Wal-Mart's aggressive and large expansion throughout the U.S., its industry leader status, and success over its common competitors such as Kmart and Target (Basker, 2007). Basker (2007) points out that the impact of Wal-Mart on local economies is quantitatively or qualitatively different from the effects of other big-box retailers such as Costco, Target, or Kmart, which still remains as an important open question (Basker, 2007).

Haltiwanger, Jarmin and Krizan (2010) find that big-box stores' entry and growth have a significant negative impact on employment growth. They use establishment-level data with detailed location information in a single metropolitan area. Effect results are especially higher on smaller chain stores when the big-box activity is in the same detailed industry, and in the immediate area. Schuetz (2015) finds that big-box stores tend to avoid the existing "own-firm" stores, and so they locate near the complementary big-box stores. She concludes that firms may prefer to share consumers in a desirable location than ceding the entire market to competitor firms.

Basker (2011) uses quarterly data from 1997-2006 to estimate the aggregate income elasticity of Wal-Mart's, and Target's, revenues. She finds that during an economic downturn Wal-Mart's revenues increase whereas Target's revenues decline. Neumark, Zhang and Ciccarella (2008) study Wal-Mart's entry effect on per capita retail (payroll) earnings and retail employment. For the periods of 1977-1995, they combine County Business Patterns data with Wal-Mart's administrative store location data. To remove endogeneity they follow the hub-and-spoke expansion pattern of Wal-Mart's store openings until 1995. This is done to isolate exogenous variations in store openings, i.e., the change in number of stores from one year to the next, by considering the inverse of the distance from Benton County, Arkansas (Wal-Mart headquarter), and yearly dummies. Their findings show that 1.4 jobs are lost for every job created by Wal-Mart in the local economy. This results in approximately 146 displaced workers or 2.7% county level reduction of the average retail employment. In Neumark, Zhang and Ciccarella (2008), Wal-Mart store openings also lead to declines in county-level retail earnings of about $1.4 million, or 1.5 percent. With a similar identification strategy, Dube, Lester and Eidlin (2007) finds that for every new Wal-Mart store opened, county-level average retail wage is reduced by 0.5% to .9%, mainly due to a decrease in labor market rents. Analogously, as in Basker (2005), both Neumark, Zhang and Ciccarella (2008) and Dube, Lester and Eidlin (2007) do not differentiate between full and par-time employment.

Drewianka and Johnson (2006) use an event analysis approach on county-level panel data. They conclude that Wal-Mart's presence increases employment in the retail sector by approximately 155 to 162 additional jobs. The number of concurrent establishments are unaffected by Wal-Mart's presence. However, some questions remain concerning their methodology. For instance, their results implicitly highlight potential reverse causality (endogeneity) on the company's location decision making. Findings in Drewianka and Johnson (2006) suggest that more Wal-Mart stores are located where employment is decreasing in relation to other considered variables. Additionally, Drewianka and Johnson (2010) do not distinguish between full-time and part-time employment. Spillover effects (spatial autocorrelation) on and between contiguous counties are disregarded as well.

Goetz and Shrestha (2009) find that Wal-Mart's presence results in higher wages relative to self-employment. They attribute this to the creative destruction process, in which smaller businesses are displaced by Wal-Mart's arrival. Results are consistent with the conclusions of other studies researching growth in labor productivity for the retail sector, during the 1990s. These studies highlight how establishments with higher productivity replaced the ones with lower productivity (Foster, Haltiwanger and Krizan, 2006). Additionally, Kolko and Neumark (2010) highlight how locally owned businesses lead to more employment stability during economic downturns. The clearest benefits do not come from small, independent businesses, but instead from corporate headquarters and, to a lesser extent, from small, locally-owned chains (Kolko and Neumark, 2010).

From a different angle, Bonanno and Lopez (2012) determine Wal-Mart's effect on workers by evaluating whether Wal-Mart exercises monopsony power over workers in the retail sector. For the contiguous states of U.S., they use county-level observations for the year 2006. Their findings reveal that Wal-Mart's potential wage reduction below the competitive level in the U.S. on average amounts to less than 3%. However, these wage reduction in non-metropolitan counties are three-fold those in metropolitan counties and are highest in non-metro areas of the south and central states but negligible in northeastern states (Bonanno and Lopez, 2012).The most relevant limitations of Bonanno and Lopez (2012) include singular focus on workers and that spillover effects from non-retailing industries are unaccounted for. Moreover, in their study retail labor is treated as homogenous, that is, there is no account for difference in skills levels and the distinction between full and part-time employment are likewise overlooked.

### 2.2. Issue of Endogeneity and Research Gap

Regarding the issue of endogeneity with respect to big-box store locations, the use of instrumental variable (IV) methods has proven inconclusive due to a lack of clear agreement. For instance, Neumark, Zhang and Ciccarella (2008) using a distance-based instruments challenge the approaches of Basker (2005), Goetz and Swaminathan (2006) and Dube, Lester and Eidlin (2007). In return, Basker (2007) discloses how the instrumental variables in Neumark et al. (2008) are likely correlated with unobservable drivers of Wal-Mart's location decision as well as with variations in business cycles. Basker (2007) argues that, in spite of the exogeneity of the distance and time variables, "…exogeneity does not automatically mean an instrument satisfies the exclusion restriction." (p.20). In other words, due to the ubiquitous presence of Wal-Mart stores across markets and counties, accounting for the distance from Bentonville (Wal-Mart headquarter) has become a rather questionable instrument.

In addition, the continuous growth of Wal-Mart and other big-box retailers like Target across the U.S. continues to be a concern for the general population, local policy makers and researchers. The past literature has been inconclusive in analyzing the dynamics[1] of the economic relationships between Wal-Mart's and Target's growth and their effect on local economies, personal income growth, and retail trade employment and wages. The literature has also missed in assessing the degree to which Wal-Mart's impact on local economies is quantitatively or qualitatively different from the effect of other "big-box" retailers such as Target (Basker, 2007).

Mixed conclusions and questionable methods have fallen short in shedding light on these issues, which is emphasized in the Bonanno and Goetz (2012) survey paper. Bonanno and Goetz (2012) mention the need for a unifying empirical framework and identification strategy to deal with the endogeneity issue of the company's store location decision - when studying its effect on local economic matters (e.g., retail employment and wages).

This dissertation estimates the impact on county retail employment and wages as a result of Wal-Mart's and Target's aggregate and individual store presence. This is done by estimating a panel VAR model to assess the dynamic endogenous relationship among the variables of interest. Subsequently, the economic impact of Wal-Mart and Target is estimated in regards to the degree in which their individual presence affect personal income growth in U.S. counties, while controlling for spatial autocorrelation. Finally, impulse response functions are calculated to evaluate the dynamic effects from the growth of big-box (Wal-Mart and Target) stores across counties relative to the personal income growth.

[1] Jantzen, et al. (2009) use cointegration techniques and causality tests to examine the relationship between Wal-Mart U.S. sales and a set of macro measures of income, employment, production, and prices. However, their study is done at the national level for the periods of 2000-2005 and 1995-2005.

# CHAPTER 3

## BIG-BOX RETAILERS, RETAIL EMPLOYMENT AND WAGES

### Abstract

This chapter applies a vector autoregression model on panel data (panel VAR) and calculates impulse response functions and variance decompositions to examine the dynamic relationships among Wal-Mart and Target stores, retail employment, and retail wages - at the county level for the period from 1986-2005. The panel VAR does not require strong assumptions which are necessary in models that make use of questionable instruments to control for endogeneity, but rather assumes all variables in the system to be endogenous to one another. Results suggest that the presence of big-box retailers does increase the number of retail jobs; however, it decreases the level of retail wages. The effect on local retail employment is relatively larger than the effect on retail wage. Wal-Mart's presence drives the results. Retail employment is a more important factor over retail wage affecting the big-box retailers' location decision.

Keywords: Panel Vector-Autoregression, Retail Employment, Retail Wage, Wal-Mart, Target.

JEL Codes: C33, J21, J23, J31, R1

**3.1. Introduction**

As Wal-Mart and Target stores continue to spread all over the U.S., there is a growing concern regarding the impact these big-box retailers have on local economic outcomes, i.e. employment, wages, poverty level, etc. As discussed in Chapter 2, numerous studies conclude positive, negative and mixed effects on the local economies where big-box retailers, especially Wal-Mart, locate their stores. Although little has been studied about the effect from Target stores, the literature on Wal-Mart is extensively surveyed in Bonanno and Goetz (2012). Bonanno and Goetz (2012) review the literature on Wal-Mart and its impacts on the local economies in detail including aspects of community life, i.e., wages and jobs in the retail sector, consumer sector, and likewise review econometric estimation issues. Bonanno and Goetz (2012) conclude that there exist positive and negative effects of Wal-Mart stores on local economies and suggest five open research questions.

One of the open questions and a main gap in the literature is "identification strategy", which is properly accounting for endogeneity among the variables considered (Bonanno and Goetz, 2012, p. 294). A main problem when endogeneity exists, is that least squares regressions tends to be biased and inconsistent (Wooldridge, 2015). The literature on Wal-Mart's economic impact is inundated with this common estimation issue. Endogeneity occurs because Wal-Mart's location decision is dependent upon the local economic indicators, for example, retail employment, wage, and poverty rate in a region (Goetz and Swaminathan, 2006). Retail wage and retail employment are determined simultaneously in the local labor market. In analogous form, as previously mentioned, Wal-Mart and Target store locations may be endogenously related to retail wage and employment level in a county. Bonanno and Lopez (2012) show that Wal-Mart's monopsony power over workers lowers retail wages.

A common claim on Wal-Mart's effect is that upon its arrival at a local community, it may eliminate more jobs than it creates while stimulating lower wages in the retail sector (Norman, 2004; Quinn, 2005, Watch, 2005). The results from Drewianka and Johnson (2006) highlight a potential reverse causality. In their findings, a greater number of Wal-Mart stores tend to locate where employment is decreasing. Similarly, retail earnings may drive Wal-Mart's location decision (Ostrander, 2011), and thus Wal-Mart appears to locate stores in areas with sequentially higher density and higher income growth. This is in line with Wal-Mart strategy of locating in small towns with increasing population growth (Slater, 2003).

This essay complements earlier work in the literature concerning the endogeneity issue among the big-box retailers' location decision, retail wage, and retail employment by using the vector autoregression model with panel data (panel VAR). The panel VAR model is useful when endogeneity and unobserved heterogeneity are present (Semykina and Wooldridge, 2010). The panel VAR approach provides a unifying empirical framework and identification strategy that has been absent in the previous literature that addressed the big-box stores' impact on the local economy. According to results from this chapter, the effect of big-box stores on retail wages is negative and significant. However, the presence of big-box stores increases retail employment. The total variation explained by the presence of the big-box stores is relatively larger in local retail employment than in retail wage. In addition, retail employment is a more important factor over retail wage, affecting the big-box stores' location decision. Interestingly, compared to Target's, Wal-Mart has larger effects given the empirical results.

In sum, this essay contributes to the regional economic literature in three ways. First, through the use of the panel VAR approach, the dynamic endogenous relationship among the variables is accounted for, allowing for county-specific unobserved heterogeneity. Second, via a reduced-form VAR model, the results shown here do not rely on strong assumptions that are necessary in models that use questionable instruments to deal with endogeneity. Third, the calculation of orthogonalized impulse response functions (IRF) and variance decompositions allows to separate the response of retail employment, wages and the big-box retailers to shocks from each of the variables of interest.

### 3.2. Data

This study uses county-level data to construct a panel with annual data for the 48 contiguous U.S. states for the period from 1986-2005. The time span of this analysis covers the periods in which many counties experienced a steady Wal-Mart expansion sprouting throughout the U.S. (Basker, 2007, Holmes, 2011). Additionally, the data is compiled up to 2005 - based on availability of county level data. Panel data allows the researcher to take advantage of both cross sectional and time series information in examining the empirical relationships among the variables. The use of panel data results in an increase in the number of observations and degrees of freedom, while also reducing any collinearity among the explanatory variables (Hurlin and Venet, 2003). In this chapter, counties are selected as the cross-sectional units for the analysis because many policies related to economic growth and development are formulated at the county level (Carlino and Mills, 1987, Deller, et al., 2001, Rupasingha, Goetz and Freshwater, 2002). The resulting sample is a panel with small variable T (time) and large variable N (counties).

The number of Wal-Mart stores during this period is compiled from Holmes (2011) database which is available at http://www.econ.umn.edu/~holmes/data/WalMart/, and normalized by population (per 100,000 inhabitants). Wal-Mart made a public file in November 2005, which lists a Wal-Mart store, address, store number, store type, and opening date. Holmes (2011) combines these data with additional information posted at the Wal-Mart website and lists the opening dates for each store. For this analysis, the count of each store county by county, and by year, is generated using the Holmes data set. The Target store count is generated using the Target store openings data from FLOWINGDATA (https://flowingdata.com/2009/10/22/target-store-openings-since-the-first-in-1962-data-now-available). A depiction of Wal-Mart and Target stores count is shown in Figure 3.1. A composite variable adding the Wal-Mart store and Target stores count is calculated for each county over time, which is then normalized by population in 100,000 - for the further use.

Retail employment data is obtained from the U.S. Census County Business Patterns (CBP). Employment and payroll data are compiled for the sectors with NAICS code 44 (retail trade sector) except 441 (motor, vehicles and parts dealer) and 447 (gasoline stations).[2] The employment data contains full-time and part-time jobs. Not having the distinction between full-time and part-time employment is a limitation in employment data given that many jobs in the retail sector may not necessarily be full-time. Employment data is normalized by population in hundreds. This gives the proportion of retail jobs per hundred residents in the county.

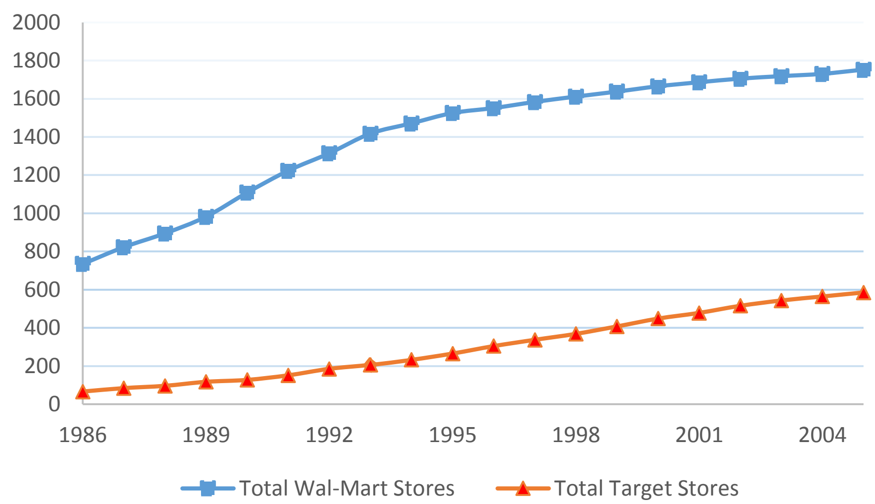


**Figure 3.1. Wal-Mart and Target store count 1986-2005**
**Source:** Produced with author's compiled data set.

[2] NAICS is the North American Industry Classification System developed by the Office of Management and Budget (OMB). NAICS 44 is the retail trade sector (http://www.census.gov/eos/www/naics/). NAICS 441 (motor, vehicle and parts dealer) and 447 (gasoline station) are excluded because they are not part of Wal-Mart's and Target's offering.

Retail payroll data is also obtained from US Census CBP. For the purpose of this study retail payroll data is utilized as a proxy of retail wage (henceforth). According to CBP, payroll includes all forms of compensation before tax, such as salaries, wages, commissions, dismissal pay, bonuses, vacation allowances, sick-leave pay, and employee contributions to qualified pension plans paid during the year to all employees. The wage data is deflated using the GDP deflator (with base year 2009), and normalized by dividing it by the number of jobs in the retail sector.

Summary statistics and variable definitions are presented in Table 3.1. The average number of Wal-Mart stores across counties is 0.77 (with a range of 0 to 30 stores across counties). Target has a much smaller average (0.21). The average number of Wal-Mart stores per 100,000 residents is 1.53. Target, on the other hand, has a considerably smaller average number of stores per 100,000 residents (0.12). The average retail jobs are about 4,109 across counties while the average retail wage is at about $17,739 (ranging from $6,985 to $69,171). The average normalized retail employment is 3.37, meaning that there are at least three persons working in retail sector per 100 people living in a county.

In preparing the sample data, 266 counties with missing retail employment and payroll information were dropped from the sample. Also, only counties with retail employment greater than 100 are included in the analysis. Dropping observations makes the panel data unbalanced. Note, however, that the results discussed in the empirical section do not significantly change if the full sample is utilized.

**Table 3.1. Summary Statistics**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Wal-Mart store count | 0.77 | 1.21 | 0 | 30 |
| Wal-Mart stores per 100,000 persons | 1.53 | 2.14 | 0 | 17.09 |
| Target store count | 0.21 | 1.03 | 0 | 43 |
| Target stores per 100,000 persons | 0.12 | 0.73 | 0 | 37.05 |
| Total retail employment (persons) | 4,109 | 12,689 | 100 | 363,073 |
| Retail employment per 100 persons | 3.38 | 1.48 | 0.49 | 23.97 |
| Total retail payroll (thousand dollars) | 67,099 | 237,296 | 533 | 7,994,974 |
| Retail wage per worker (thousand dollars) | 17.74 | 3.35 | 6.98 | 69.17 |
| Big-box stores count (Wal-Mart + Target stores) | 0.98 | 1.99 | 0 | 68 |
| Big-box store per 100,000 persons | 1.65 | 2.26 | 0 | 43.22 |
| County population (persons) | 95,390 | 296,130 | 1,587 | 9,793,263 |
| N = 54,242 | | | | |
| n = 2,882 and T = 19* | | | | |

*Note*: 1. Wal-Mart data are compiled from Holmes (2011) database which is available at http://www.econ.umn.edu/~holmes/data/WalMart/
2. Target store data are compiled from https://flowingdata.com/2009/10/22/target-store-openings-since-the-first-in-1962-data-now-available/
3. Retail employment and payroll data are compiled from the U.S. Census County Business Patterns (CBP)
* For some counties, T < 19, which makes the data unbalanced

### 3.3. Panel Vector Autoregressions (VAR)

This chapter analyzes the dynamic inter-relationship among the county's presence of Wal-Mart and Target, retail employment, and retail wage. It also studies how these relationships may determine the companies' location decision. A panel VAR model developed by Holtz-Eakin, Newey, and Rosen (1988) and Love and Zicchino (2006) is applied to the data. The dynamic effects among the relationships are explained graphically using the orthogonalized impulse response functions (IRFs).

Variance decompositions are also reported, which show the percent variation in one variable that is explained by the shock or innovation to another variable accumulated over time (Hamilton, 1994). In other words, the variance decompositions give the magnitude of the total effect. As said, the panel VAR provides a unifying empirical framework and identification strategy that has been absent in the previous Wal-Mart literature.

The variables of interest are retail wage, retail employment and the (previously normalized) number of Wal-Mart and Target stores. These variables are endogenous to one another. Retail wage and retail employment are determined simultaneously in the local labor market. Similarly, Wal-Mart and Target store location may be endogenously related to retail wage and employment at the county level. For example, Bonanno and Lopez (2012) show that Wal-Mart's monopsony power over workers lowers retail wages.

A common claim on Wal-Mart's effect is that upon its arrival at a local community, it may eliminate more jobs than it creates while encouraging lower wages in the retail sector (Norman, 2004; Quinn, 2005, Watch, 2005). The results in Drewianka and Johnson (2006) highlight the potential reverse in causality. In their findings a greater number of Wal-Mart stores tend to locate where employment is decreasing. In addition, Ostrander (2011) finds significant correlation between Wal-Mart stores location decision and the local level of population density and average household income. This goes along with Wal-Mart strategy of locating in small towns with increasing population growth (Slater, 2003).

The panel VAR methodology fits the objective of this study since there is no a priori theory concerning the relationship among Wal-Mart's and Target's presence, retail employment and wages. The panel VAR is the technique that combines the traditional VAR approach, for which all variables in the system are considered endogenous, and panel data allowing for unobserved individual heterogeneity (Love and Ariss, 2014, Love and Zicchino, 2006). Likewise, the framework of the panel VAR allows for endogenous relationships among the variables that enter a system of equations, within which the short-run dynamic relationships may be later identified (Koutsomanoli-Filippaki and Mamatzakis, 2009).

Moreover, this methodology facilitates the isolation of the responses of retail employment and wages to shocks from Wal-Mart's and Target's presence, through calculation of IRFs. The orthogonalized IRF shows the reaction of one variable of interest (i.e., retail employment) to a shock in another variable of interest (i.e., number of Wal-Mart and Target stores). Therefore, the orthogonalization of each response allows the identification of the impact of one shock at a time while keeping all other shocks constant (Hamilton, 1994).

The general panel VAR takes the following reduced form:

$$\boldsymbol{Y}_{it} = \boldsymbol{\theta}(L)\boldsymbol{Y}_{it} + \boldsymbol{f}_i + \boldsymbol{\tau}_t + \boldsymbol{\varepsilon}_{it}, \tag{3-1}$$

where $\boldsymbol{Y}_{it} = [y_{it}^1, y_{it}^2, \dots, y_{it}^p]'$ is a vector containing the variables of interest. The $i$ subscript denotes county while the t subscript represents time period. $\boldsymbol{\theta}(L)$ is a polynomial matrix in the lag operator, that is, $\boldsymbol{\theta}(L) = \boldsymbol{\theta}_1 L^1 + \boldsymbol{\theta}_2 L^2 + \cdots + \boldsymbol{\theta}_k L^k$. A time-invariant region-specific element ($\boldsymbol{f}_i$) is included to control for county-specific effects that may be unobserved or omitted heterogeneity (e.g. geographical location, climate, amenities, land-use policy, etc.). Similarly, a county-invariant time-specific element $\boldsymbol{\tau}_t$ is included to account for possible shocks common across counties but varying over time (e.g. business cycle effects, fiscal policies, and technological progress). Lastly, $\boldsymbol{\varepsilon}_{it} \sim i.i.d.\ N(\boldsymbol{0}, \Omega)$ is a vector of idiosyncratic errors.

In order to use the VAR procedure with panel data, the same underlying structure for each cross-sectional unit is needed (Love and Zicchino, 2006). However, in practice, this constraint is usually violated. As a way to overcome this restriction on parameters, "individual heterogeneity" is allowed in the variables' levels. This is given by the fixed effects $\mathbf{f}_\mathrm{i}$ in the model. Unfortunately, the fixed-effects estimator is not consistent in a dynamic panel. In other words, the fixed effects due to lags of the dependent variables are correlated with the regressors (Nickell, 1981).

As a result, the commonly used mean-differencing procedure to eliminate fixed effects would generate biased coefficients. To address this, following Arellano and Bover (1995) and Love and Zicchino (2006), application of the forward mean differencing - also known as the Helmert Transformation procedure - is used to eliminate the fixed effects. In this procedure the forward mean (the mean of all available future observations for each county-year) is removed. The time fixed effects, $\boldsymbol{\tau}_t$, which is included in equation (3-1) to control for macro shocks, is also eliminated during the Helmert transformation.

The orthogonality between transformed variables and lagged regressors is preserved. Therefore, lagged regressors can be used as instruments and equation (3-1) can be estimated via the system generalized method of moment (GMM) as explained in Arellano and Bover (1995). Since the model in equation (3-1) is just identified, (i.e., the number of instruments equals the number of regressors) it may also be estimated using two stage least squares (2SLS).

### *3.3.1. The Helmert Transformation*

Let $y_{it}^p$ denote a variable in the vector $\boldsymbol{Y}_{it}$. Define

(3-2) $$\bar{y}_{it}^p = \frac{1}{T}\sum_{t=1}^{T} y_{it}^p,$$

where $\bar{y}_{it}^p$ denotes the mean of $y_{it}^p$ over time for the $i$th county. Then,

(3-3) $$\tilde{y}_{it}^p = \frac{1}{T-t}\sum_{n=t+1}^{T} y_{in}^p,$$

where $\tilde{y}_{it}^p$ represents the means obtained from the future values of $y_{it}^p$, where $T$ denotes the last period of data for a given series. Let

(3-4) $$\tilde{\varepsilon}_{it}^p = \frac{1}{T-t}\sum_{n=t+1}^{T} \varepsilon_{it}^p,$$

where $\tilde{\varepsilon}_{it}^p$ denotes a similar transformation for $\varepsilon_{it}^p$; and $\varepsilon_{it} = \left[\varepsilon_{it}^1, \varepsilon_{it}^2, \dots, \varepsilon_{it}^p\right]'$. Thus, the Helmert-transformed versions of $y_{it}^p$ and $\varepsilon_{it}^p$ are given by:

(3-5) $$\hat{y}_{it}^p = \lambda_{it}(y_{it}^p - \tilde{y}_{it}^p),$$

and

(3-6) $$\hat{\varepsilon}_{it}^p = \lambda_{it}(\varepsilon_{it}^p - \tilde{\varepsilon}_{it}^p),$$

where $\lambda_{it} = \sqrt{(T-t)/(T-t+1)}$.

As implied by equations (3-5) and (3-6), the Helmert observation for time t is the difference between the observation for time t and the observations at time $t+1$ through $T$. That is, the mean of all future observations. Note that for the last year of data available, the Helmert transformation cannot be computed. This is because there is no future value for the forward mean construction.

### *3.3.2. Empirical Model*

The transformed model using the Helmert procedure is given by:

(3-7) $$\widehat{\boldsymbol{Y}}_{it} = \boldsymbol{\theta}(L)\widehat{\boldsymbol{Y}}_{it} + \hat{\boldsymbol{\varepsilon}}_{it},$$

where $\widehat{\boldsymbol{Y}}_{it} = \left(\hat{y}_{it}^{1}, \hat{y}_{it}^{2}, \dots, \hat{y}_{it}^{p}\right)'$and $\hat{\boldsymbol{\varepsilon}}_{it} = \left(\hat{\varepsilon}_{it}^{1}, \hat{\varepsilon}_{it}^{2}, \dots, \hat{\varepsilon}_{it}^{p}\right)'$. In this transformation all observations are expressed as deviations from average future observations. In (3-7) the Helmert transformation gives larger weight to the observations that are closer to the beginning of the time series. Additionally, if the errors are not autocorrelated before the transformation, similar properties should hold afterwards. In other words, the transformation does not induce serial correlation and likewise preserves homoscedasticity (Arellano and Bover, 1995). As mentioned above, this technique allows using lagged value of regressors as instruments, and the utilization of GMM to estimate the coefficients.

IRFs are generated from the estimation of all the coefficients of the panel VAR in equation (3-7). As discussed earlier, IRF describes how endogenous variables respond to a shock in another variable in the system, while holding all others constant. Following Love and Zicchino (2006), confidence intervals for the IRF are computed with Monte-Carlo simulations. The coefficients in equation (3-7), their respective variance-covariance matrix, and IRFs are drawn randomly. This procedure is repeated 1000 times. This allows building a distribution with its 5th and 95th percentiles. Nonetheless, it is unlikely that the variance-covariance matrix of the errors will be diagonal.

Therefore, to separate the shocks applied to each one of the system's variables, the residuals need to be decomposed in a manner that they become orthogonal. As a convention, a particular ordering is adopted, and any correlation between the residuals of any two elements is allocated to the variable that appears first in the ordering. The Cholesky decomposition is then used to compute the IRF.

This decomposition assumes that series that come earlier in the ordering have a contemporaneous effect on the following variables, as well as through a lag (Hamilton, 1994). The variables that enter after will only affect the previous variables through a lag. More specifically, earlier series in the system are considered more exogenous than the ones that appear later. These latter, in turn, are considered more endogenous. For the model in (3-7), the variable ordering choice can be made on the basis of a priori knowledge on the structure of the relationships between the system's variables.

Recall the variables of interest. They are retail wage ($rw_{it}$) per worker, (normalized) retail employment ($re_{it}$), and (normalized) number of the big-box stores ($bx_{it}$) by county. As discussed above, these variables all possess an endogenous relationship. By construction, at the equilibrium, retail wage and employment are determined simultaneously and therefore ordering should not matter. However, for this study, the big-box variable is always assumed to be the most endogenous variable and thus comes last in the ordering. This assumption implies that the effect of the presence of the big-box stores on retail wage and employment may take effect with at least 1-year lag. For this study we run the baseline model with the following ordering:

(3-8) $$(rw_{it}, re_{it}, bx_{it}).$$

The ordering in equations (3-8) gives rise to the following system of equations assuming the optimal lag is one, i.e., panel VAR (1):

(3-9) $$rw_{it} = \alpha_{10} + \alpha_{11} rw_{i,t-1} + \alpha_{12} re_{i,t-1} + \alpha_{13} bx_{i,t-1} + \varepsilon_{it}^{rw}$$

$$re_{it} = \alpha_{20} + \alpha_{21} rw_{i,t-1} + \alpha_{22} re_{i,t-1} + \alpha_{23} bx_{i,t-1} + \phi_{21} rw_{it} + \varepsilon_{it}^{re}$$

$$bx_{it} = \alpha_{30} + \alpha_{31} rw_{i,t-1} + \alpha_{32} re_{i,t-1} + \alpha_{33} bx_{i,t-1} + \phi_{31} rw_{it} + \phi_{32} re_{it} + \varepsilon_{it}^{bx}$$

The ordering sequence in (3-8), which corresponds to the system in equation (3-9), shows that retail wage comes first meaning that it is more exogenous, having a contemporaneous effect and also with a lag on retail employment and the number of the big-box stores. Similarly, the ordering in equation (3-8) implies that retail employment affects the number of the big-box stores contemporaneously and also with a lag, while it affects retail wage only with a lag. In turn, the big-box stores affect retail wage and retail employment only with a lag. In the results section, the alternative ordering in (3-10) is also reported. This ordering separates the effect of Wal-Mart and Target stores presence:

(3-10) $$(rw_{it}, re_{it}, wm_{it}, tg_{it}),$$

where $\mathrm{wm}_{\mathrm{it}}$ is (normalized) number of Wal-Mart stores and $\mathrm{tg}_{\mathrm{it}}$ is (normalized) number of Target stores.

## 3.4. Empirical Results

### *3.4.1. Panel Unit Root Test*

To use the panel VAR approach, all the variables need to be stationary. Hence, testing the unit root is the first phase of the analysis. There are two classes of tests that can be used to detect the presence of the unit roots in the panel data. The first-generation panel unit root tests by Hadri (2000) has been developed assuming cross-section independence across units in the panel (with the exception of common time effects). In second-generation tests, the assumption of cross-sectional independence is relaxed, which allows for an array of dependence among the different units (Pesaran, 2007, Smith, et al., 2004).

To check for the presence of unit roots in the series, the Fisher's test as suggested in Maddala and Wu (1999) is used. The Fisher's test allows for heterogeneity in the autoregressive coefficient of the Dickey-Fuller regression and ignores cross-sectional dependence in the data. The test combines the p-values from N independent unit root tests (Maddala and Wu, 1999). In addition, the test does not require a balanced panel and allows the existence of gaps (different time spans across cross-sectional units). This is convenient since the sample data consists of an unbalanced panel with gaps. Table 3.2 shows the results for the unit root tests.

The null hypothesis for both tests is that all series are non-stationary while the alternative is that at least one of the series is stationary. The results suggest that retail wage and employment are stationary in level. The big-box store variable, however, is integrated of order one. Similarly, the number of Wal-Mart and Target store series are non-stationary in levels and are integrated of order one.

**Table 3.2. Fisher Panel Unit Root Tests**

| Variables | Augmented Dickey-Fuller | Phillips-Perron |
|---|---|---|
| $rw_{it}$ (Retail wage rate) | | |
| Level | 9526*** | 15000*** |
| Difference | 24100*** | 61300*** |
| $re_{it}$ (Retail employment) | | |
| Level | 8441*** | 11900*** |
| Difference | 23200*** | 52400*** |
| $bx_{it}$ (Normalized number of big-box stores) | | |
| Level | 3739 | 3605 |
| Difference | 8409*** | 18000*** |
| $wm_{it}$ (Normalized number of Wal-Mart stores) | | |
| Level | 4155 | 3819 |
| Difference | 7940*** | 16800*** |
| $tg_{it}$ (Normalized number of Target stores) | | |
| Level | 831 | 784 |
| Difference | 2926 | 6523*** |

*Note:* All unit root tests are performed with 1 lag and a trend. (*), (**), (***) represents significance at the 10%, 5% and 1% level respectively.

### *3.4.2. Panel VAR Estimation Results*

The panel VAR in equation (3-7) is estimated using a PVAR package in Stata developed by Love and Zicchino (2006). The impact of the presence of these big-box stores on retail wage and employment is assessed with the variable sequence ordering in (3-8). Results for the alternative sequence ordering (3-10) are shown and discussed subsequently. The optimal lag length for the panel VAR model in (3-7) is selected based on the moment model selection criteria (MMSC) developed by Andrews and Lu (2001).

Table 3.3 reports the MMSC Bayesian information criterion (MBIC), MMSC Akaike's information criterion (MAIC), and MMSC Hannan and Quinn information criterion (MQIC). Similar to maximum likelihood-based information criteria (i.e. AIC, BIC and HQIC), the model which minimizes the MAIC, MBIC or MQIC is the preferred model. Consequently, for the model in (3-7) e optimal lag length is 3. Note that for just identified systems like in (3-7), the Hansen's (1982) J statistic is equal to the MAIC, MBIC and MQIC.

**Table 3.3. Lag Order of Panel VAR Selection**

| Lag | CD | J | J *pvalue* | MBIC | MAIC | MQIC |
|---|---|---|---|---|---|---|
| 1 | .9838 | 1.81e-29 | 0.00 | 1.81e-29 | 1.81e-29 | 1.81e-29 |
| 2 | .9879 | 2.48e-29 | 0.00 | 2.48e-29 | 2.48e-29 | 2.48e-29 |
| 3 | .9904 | 6.87e-30 | 0.00 | 6.87e-30* | 6.87e-30* | 6.87e-30* |
| 4 | .9920 | 2.17e-29 | 0.00 | 2.17e-29 | 2.17e-29 | 2.17e-29 |
| No. of obs. | 44,961 | | | | | |
| No. of panels | 2,785 | | | | | |
| Average no. of T | 16.14 | | | Sample: | 1988 - 2004 | |

*Note:* These statistics are produced using the *pvarsoc*, a Stata module which reports the coefficient of determination (CD), J statistics as in Hansen (1982) and corresponding *p-value* (J *pvalue*). Also this table reports moment model selection criteria developed by Andrews and Lu (2001): MMSC-Bayesian information criterion (MBIC), MMSC-Akaike's information criterion (MAIC), and MMSC-Hannan and Quinn information criterion (MQIC).

The estimation results of the panel VAR in equation (3.7) are presented in Table 3.4. The results show that retail wage per worker ($rw_{it}$) responds negatively in all three lags to the change in the number of the big-box stores ($\Delta bx_{it}$). These results are statistically significant and are in line with Dube, Lester and Eidlin (2007), Neumark, Zhang and Ciccarella (2008), Bonanno and Lopez (2012) which report declines in retail wages as a result of Wal-Mart stores' openings. Retail employment, the number of jobs per 100 residents in a county, has a positive and significant effect on retail wage. This is consistent with economic theory regarding the relationship between labor supply-demand. Similarly, retail wage has a positive and significant impact on retail employment for the first lag.

The impact of the change in the number of the big-box stores ($\Delta bx_{it}$) on retail employment ($re_{it}$) is positive and significant for the first lag.[3] This result is consistent with Drewianka and Johnson (2006), in which Wal-Mart's presence marginally increases employment in the retail sector of the local economy. In addition, Hicks and Wilburn (2001), Basker (2005) and Hicks (2007) also finds a moderate creation of jobs as a result of Wal-Mart's presence.

Drewianka and Johnson (2006) underline the potential reverse in causality between Wal-Mart's location decision and local retail wage and employment. This is consistent with the results shown in Table 3.4, where a higher level of retail wage ($rw_{it}$) results in a decrease in the change of the big box stores ($\Delta bx_{it}$). This highlights the big-box retailers' preference to locate in counties with rather lower retail wages. In turn, a higher level of employment in the first lag draws in a higher change in the number of big-box stores, but the coefficient is not significant at any conventional significance levels.

[3] The exact source for the gains in jobs in the retail sector is unknown. However, some complementary establishments may yield positive benefits from Wal-Mart's presence , which may have an impact on retail job gains .

**Table 3.4. Panel VAR Estimation Results ($\boldsymbol{rw_{it}}, \boldsymbol{re_{it}}, \boldsymbol{bx_{it}}$)**

| Response of | | Response to $rw_{it}$ | $re_{it}$ | $\Delta bx_{it}$ |
|---|---|---|---|---|
| $rw_{it}$ (Retail wage rate) | | | | |
| | L1. | 0.6345 | 0.4007 | -0.0975 |
| | | [32.70]*** | [4.00]*** | [-3.49]*** |
| | L2. | 0.1678 | 0.0597 | -0.0361 |
| | | [10.46]*** | [0.83] | [-2.05]** |
| | L3. | 0.0786 | 0.0168 | -0.0355 |
| | | [8.73]*** | [0.41] | [-2.57]*** |
| $re_{it}$ (Retail employment) | | | | |
| | L1. | 0.0235 | 0.6537 | 0.0312 |
| | | [7.17]*** | [14.17]*** | [4.28]*** |
| | L2. | -0.0127 | 0.0492 | -0.0021 |
| | | [-4.23]*** | [2.05]** | [-0.39] |
| | L3. | -0.0131 | 0.0144 | -0.0005 |
| | | [-5.69]*** | [0.93] | [-0.15] |
| $\Delta bx_{it}$ (Big-box stores) | | | | |
| | L1. | -0.0082 | 0.0266 | 0.0027 |
| | | [-3.39]*** | [1.19] | [0.48] |
| | L2. | 0.0005 | 0.0074 | -0.0071 |
| | | [0.23] | [0.65] | [-2.87]*** |
| | L3. | -0.0003 | 0.0069 | 0.0012 |
| | | [-0.18] | [0.90] | [0.28] |
| *No. of obs.* | 39,146 | | | |
| *No. of panels* | 2744 | | | |

*Note:* $\Delta \mathrm{bx}_{\mathrm{it}}$ is the number of big-box stores in first difference. The three-variable panel VAR model is estimated by GMM, fixed effects are removed prior to estimation (see Methodology section for more details). Reported numbers show the coefficients of regressing the row variables on 3 lags of the column variables. Lag selection criteria follows the model selection criteria in Table 3.3. T-statistics are in brackets. ***, ** and * indicates significance at 1%, 5% and 10%, respectively.

Given the dominance of Wal-Mart's presence over Target as discussed in Basker (2007) and as clearly seen in Figure 3.1, one must ask whether Wal-Mart's store count is driving the results in Table 3.4. To disentangle the "big-box effect", the model in (3-7) is also estimated for the alternative variable sequence ordering in (3-10). In this model, panel VAR(2)[4], Wal-Mart and Target stores' presence are assessed separately. The results in Table 3.5 reveal Wal-Mart's driving force in the previous results.

[4] The optimal lag of 2 for the sequence ordering in (3-10) was determined in a similar fashion (not shown here) as it was done for (3-8).

**Table 3.5. Panel VAR Estimation Results ($rw_{it}, re_{it}, wm_{it}, tg_{it}$)**

| Response of | | Response to $rw_{it}$ | $re_{it}$ | $\Delta wm_{it}$ | $\Delta tg_{it}$ |
|---|---|---|---|---|---|
| $rw_{it}$ | | | | | |
| | L1. | 0.6871 | 0.5142 | -0.1210 | 0.0760 |
| | | [35.97]*** | [4.57]*** | [-6.33]*** | [0.61] |
| | L2. | 0.2254 | 0.1839 | -0.0367 | -0.0304 |
| | | [12.12]*** | [2.33]** | [-2.23]** | [-0.58] |
| $re_{it}$ | | | | | |
| | L1. | 0.0170 | 0.6332 | 0.0366 | 0.0061 |
| | | [5.50]*** | [12.65]*** | [4.99]*** | [0.66] |
| | L2. | -0.0231 | 0.0542 | 0.0036 | -0.0040 |
| | | [-8.28]*** | [2.02]** | [0.65] | [-0.63] |
| $\Delta wm_{it}$ | | | | | |
| | L1. | -0.0102 | 0.0149 | -0.0010 | -0.0031 |
| | | [-4.48]*** | [0.82] | [-0.48] | [-0.91] |
| | L2. | -0.0011 | 0.0088 | -0.0021 | -0.0048 |
| | | [-0.66] | [0.78] | [-1.41] | [-1.56] |
| $\Delta tg_{it}$ | | | | | |
| | L1. | -0.0003 | -0.0024 | 0.0002 | 0.0178 |
| | | [-0.27] | [-0.29] | [0.11] | [0.63] |
| | L2. | 0.0011 | 0.0002 | -0.0004 | -0.0234 |
| | | [1.26] | [0.04] | [-0.41] | [-1.70] |
| *No. of obs.* | 42024 | | | | |
| *No. of panels* | 2765 | | | | |

*Note:* $\mathrm{rw_{it}}, \mathrm{re_{it}}, \Delta\mathrm{wm_{it}}, \Delta\mathrm{tg_{it}}$ are retail wage rate, retail employment, number of Wal-Mart stores (in first difference), number of Target stores (in first difference), respectively at the county level. The four-variable VAR model is estimated by GMM, fixed effects are removed prior to estimation (see Methodology section for more details). Reported numbers show the coefficients of regressing the row variables on 2 lags of the column variables. Lag selection criteria (not shown in Appendix) follows same approach as the one used for variable ordering (3-4). T-statistics are in brackets. ***, ** and * indicates significance at 1%, 5% and 10%, respectively. Notice the Wal-Mart and Target variables enter the model in first difference as they are both integrated of order one.

For the alternative ordering, the panel VAR estimates show a negative and significant impact on retail wages as a result of the change in the number of Wal-Mart stores ($\Delta\mathrm{wm_{it}}$). Meanwhile, Target's presence alone appears not to have a significant impact on retail wages and jobs. Conversely, Wal-Mart's presence yields a positive and significant effect on retail jobs. While the level of retail wage and employment have no impact on Target's location decision, higher retail wages are detrimental to Wal-Mart's location decision. This implies that Wal-Mart may have a preference to locate in counties with rather lower retail wages.

### *3.4.3. Impulse Response Functions*

For an improved assessment of the dynamics of the estimated effects in Tables 3.4 and 3.5, the IRFs are generated and presented in Tables 3.6 and 3.7, and Figure 3.2 and 3.3, respectively. Due to the possible contemporaneous correlations between errors, the orthogonalized IRF are used. The orthogonalization of each response allows the identification of the impact of one shock at a time while keeping all other shocks constant. Tables 3.6 and 3.7 show the details of the corresponding impulse-response magnitudes.

**Table 3.6. Impulse Response Magnitudes ($\boldsymbol{rw_{it}, re_{it}, bx_{it}}$)**

| | Time | *rw* | *re* | $\Delta bx$ |
|---|---|---|---|---|
| *rw* | 0 | 1.5001 | 0.0000 | 0.0000 |
| | 1 | 0.9392 | 0.1324 | -0.0431 |
| | 2 | 0.8536 | 0.1913 | -0.0379 |
| | 3 | 0.8233 | 0.2266 | -0.0431 |
| | 4 | 0.7406 | 0.2496 | -0.0346 |
| | 8 | 0.5056 | 0.2459 | -0.0211 |
| | 12 | 0.3387 | 0.1914 | -0.0132 |
| | 16 | 0.2247 | 0.1365 | -0.0084 |
| *re* | 0 | -0.0276 | 0.3389 | 0.0000 |
| | 1 | 0.0176 | 0.2226 | 0.0138 |
| | 2 | 0.0127 | 0.1655 | 0.0071 |
| | 3 | -0.0030 | 0.1270 | 0.0047 |
| | 4 | -0.0052 | 0.0957 | 0.0037 |
| | 8 | -0.0115 | 0.0287 | 0.0017 |
| | 12 | -0.0106 | 0.0064 | 0.0009 |
| | 16 | -0.0080 | -0.0004 | 0.0005 |
| $\Delta bx$ | 0 | 0.0159 | 0.0349 | 0.4421 |
| | 1 | -0.0130 | 0.0091 | 0.0012 |
| | 2 | -0.0068 | 0.0071 | -0.0024 |
| | 3 | -0.0066 | 0.0069 | 0.0011 |
| | 4 | -0.0064 | 0.0043 | 0.0006 |
| | 8 | -0.0049 | -0.0001 | 0.0003 |
| | 12 | -0.0034 | -0.0011 | 0.0002 |
| | 16 | -0.0024 | -0.0011 | 0.0001 |

*Note:* All variables are included in levels except for big-box that is included in differences. Each cell shows a response of the row variable to a shock in column variable (at a given time).

Each one of the three variables in the model receives a shock equal to one standard deviation of its residual while holding all other variables' innovations constant. The IRF graphs show how each variable responds to such a shock. The vertical axis shows the direction and size of the shock. The x-axis indicates the time period elapsed in years after the shock is given. The dashed lines represent a $\pm 2$ standard error confidence bound for variables' responses. As a result of having three equations in the three-variable VAR model, there are 9 IRFs.

As depicted in Figure 3.2 a one standard deviation shock to the change in big-box stores ($\Delta bx_{it}$) has a negative and significant impact on retail wage ($rw_{it}$), as shown in the upper right corner of Figure 3.2. The negative effect peaks during the fourth period and gradually declines for the remaining time horizon.

Moreover, in Figure 3.2 retail wages show a positive and significant response to a shock in retail employment, which is in line with the estimated coefficient in Table 3.4. This is also consistent with economic theory regarding the relationship between labor supply-demand. On the other hand, retail employment has an initial negative and significant response to a shock in retail wages, which follows the income effect theory (e.g. with higher wages workers would allocate more time to leisure resulting in lower number of jobs in the retail sector). However, as seen in Figure 3.2, the substitution effect takes over somewhere after the first period. This is given by the positive response to the shock after the first year. The positive response peaks after the second period when it decreases and later becomes negative.

These responses to the shock in wages are analogous to the income and substitution effect in economic theory. In addition, a one standard deviation shock to the change in big-box stores has a positive and significant effect on retail employment. The maximum effect after the shock is experienced during the first period. This positive effect, however, gradually dissipates shortly after the fifth period. This is similar to Basker (2005), Hicks and Wilburn (2001) findings, in which Wal-Mart's presence positive effect on job creation quickly vanishes after the sixth year. This latter comparison is valid, provided that Wal-Mart's presence is leading the effect of the big-box retailers' presence.

As opposed to the estimated results in Table 3.4, the IRF graph (bottom left corner of Figure 3.2) shows how the change in the number of big-box stores initially responds positively to a shock in retail wage. The effect is statistically significant, although it then turns negative and negatively peaks at the first period. This result implies that big-box stores like Wal-Mart and Target may initially be lured into a high growth area in terms of wages.

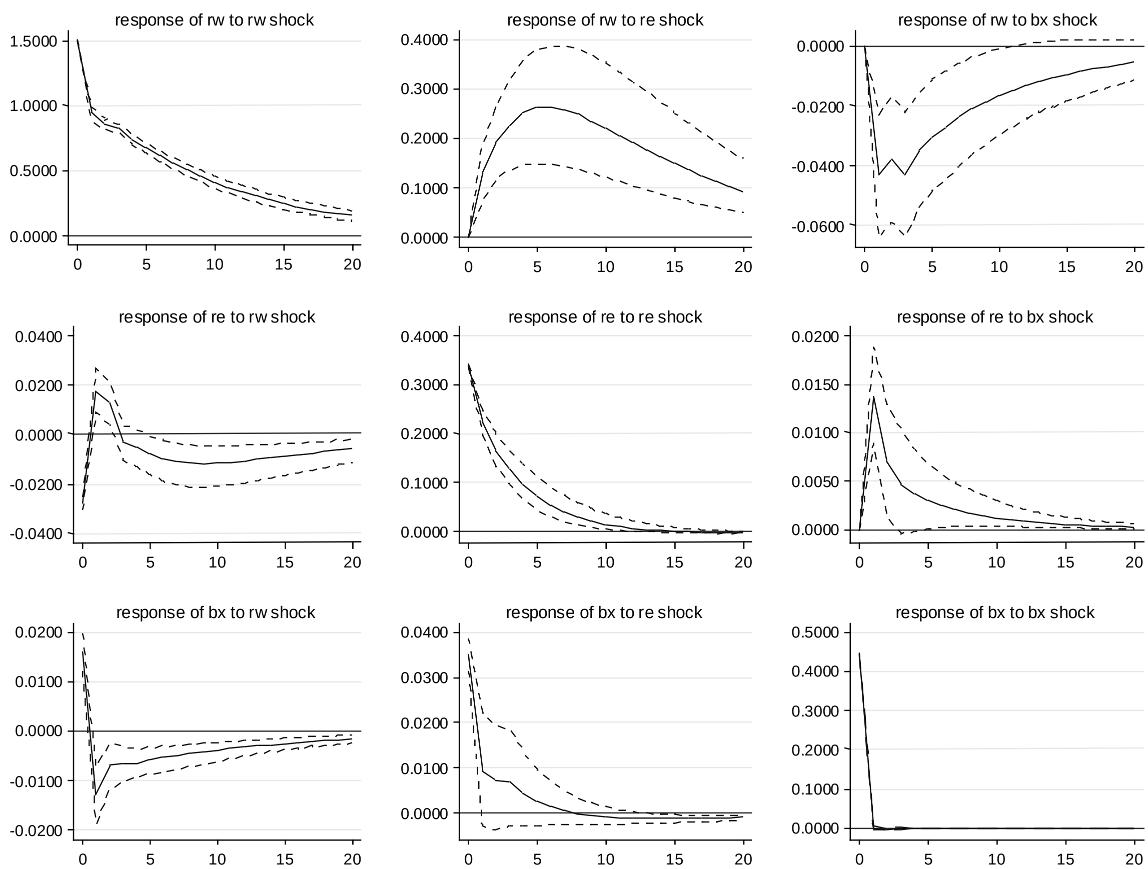


**Figure 3.2. Impulse response functions**

Note: Every row presents the different shocks to retail wage (rw), retail employment (re), and change in the number of big box stores, respectively. Errors are 5% on each side generated by Monte-Carlo with 1000 repetitions.

However, this response reverses after the first period, and thus shows a negative response to the initial shock in retail wage for the rest of the time horizon. On the other hand, a shock in retail employment drives an upward change in the big-box stores. The effect, although positive, becomes statistically insignificant after the second period. This implies that, in aggregate, Wal-Mart and Target will locate in areas with positive employment growth at least in the short run.

Figure 3.3 shows the individual effect from Wal-Mart and Target on retail wage and employment, after applying one standard deviation shock. The impulse-response magnitudes are shown in Table 3.7. The response in retail wage relative to a shock in the change of Wal-Mart stores is almost identical to that of the aggregate effect of big box stores in Figure 3.2 - implying that Wal-Mart is, in fact, leading the aggregate effect. Target's individual shock has a small positive but rather insignificant effect on retail wage.[5] This once more supports the claim that Wal-Mart seems to be leading the negative effect on retail wage, as shown in Figure 3.2. Additionally, from Figure 3.3, a shock to Wal-Mart's presence has an almost identical effect on retail jobs as shown in Figure 3.2. Target's individual effect on employment is likewise positive but rather small and statistically insignificant.

As shown in Figure 3.3, the individual response of Wal-Mart's store location decision to a shock in retail wage follows a similar pattern to that of the aggregate big-box stores. That is, the change in Wal-Mart's number of stores responds positively to one standard deviation shock in retail wage. The effect is statistically significant but it turns negative during the first period. These results attest to the endogeneity issue of Wal-Mart location and retail wages, and is in line with Ostrander (2011), which concludes that Wal-Mart seems more likely to locate in higher density and higher income localities. However, this strategy changes after the first period as shown by the impulse response, and thus supporting the claim that Wal-Mart locates in rather lower retail wage localities.

[5] Note that firms with labor unions generally have no impact on wages. This may be the reason why Target's effect on retail wage is statistically insignificant.

Target's individual response to a shock in retail wages is positive, which corroborates with Target focusing on a more affluent audience as in Basker (2007). Nevertheless, this effect is insignificant after the first period. The individual response for Wal-Mart and Target to a shock in retail employment are also shown in Figure 3.3. For both Wal-Mart and Target, a shock in retail employment results in an upward change in the number of stores. However, the effect on Wal-Mart turns insignificant after the second period while for Target it is insignificant for the entire horizon.

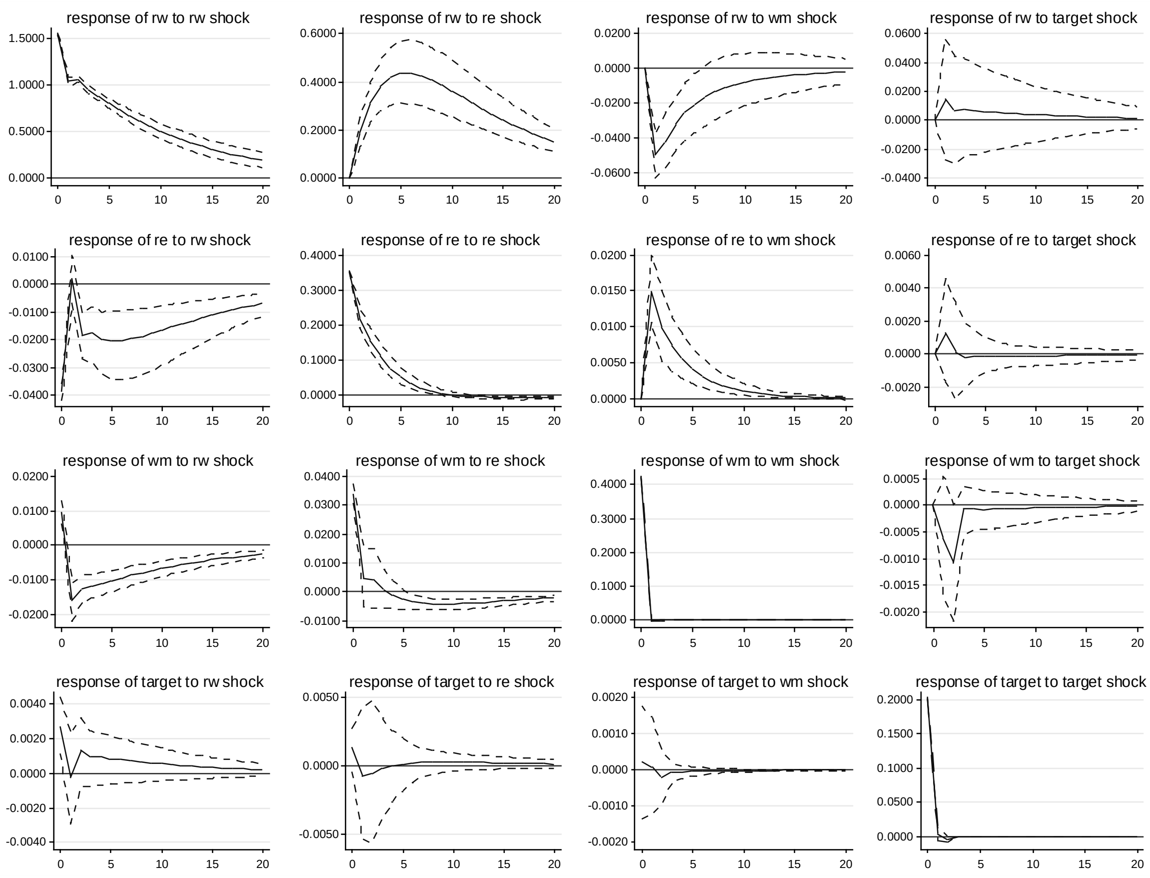


**Figure 3.3. Impulse response functions with alternative ordering**

Note: Every row presents the different shocks to retail wage (rw), retail employment (re), change in the number of Wal-Mart stores ($\Delta$wm), and change in the number of Target stores ($\Delta$tg), respectively. Errors are 5% on each side generated by Monte-Carlo with 1000 repetitions

**Table 3.7. Impulse Response Magnitudes ($rw_{it}, re_{it}, wm_{it}, tg_{it}$)**

| | Time | $rw$ | $re$ | $\Delta wm$ | $\Delta tg$ |
|---|---|---|---|---|---|
| $rw$ | 0 | 1.5407 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 1.0384 | 0.1904 | -0.0498 | 0.0149 |
| | 2 | 1.0522 | 0.3130 | -0.0420 | 0.0063 |
| | 3 | 0.9470 | 0.3831 | -0.0318 | 0.0076 |
| | 4 | 0.8743 | 0.4217 | -0.0255 | 0.0067 |
| | 8 | 0.6087 | 0.4075 | -0.0112 | 0.0047 |
| | 12 | 0.4176 | 0.3119 | -0.0059 | 0.0032 |
| $re$ | 0 | -0.0385 | 0.3513 | 0.0000 | 0.0000 |
| | 1 | 0.0020 | 0.2189 | 0.0150 | 0.0013 |
| | 2 | -0.0183 | 0.1572 | 0.0098 | 0.0001 |
| | 3 | -0.0174 | 0.1098 | 0.0073 | -0.0002 |
| | 4 | -0.0199 | 0.0756 | 0.0055 | -0.0001 |
| | 8 | -0.0187 | 0.0117 | 0.0017 | -0.0001 |
| | 12 | -0.0142 | -0.0036 | 0.0006 | -0.0001 |
| $\Delta wm$ | 0 | 0.0096 | 0.0337 | 0.4216 | 0.0000 |
| | 1 | -0.0160 | 0.0049 | -0.0004 | -0.0006 |
| | 2 | -0.0125 | 0.0041 | -0.0002 | -0.0011 |
| | 3 | -0.0119 | 0.0007 | 0.0007 | -0.0001 |
| | 4 | -0.0111 | -0.0013 | 0.0006 | -0.0001 |
| | 8 | -0.0079 | -0.0042 | 0.0002 | -0.0001 |
| | 12 | -0.0055 | -0.0038 | 0.0001 | 0.0000 |
| $\Delta tg$ | 0 | 0.0027 | 0.0013 | 0.0002 | 0.2019 |
| | 1 | -0.0002 | -0.0007 | 0.0001 | 0.0035 |
| | 2 | 0.0013 | -0.0005 | -0.0002 | -0.0045 |
| | 3 | 0.0009 | -0.0002 | -0.0001 | -0.0001 |
| | 4 | 0.0009 | 0.0000 | 0.0000 | 0.0001 |
| | 8 | 0.0007 | 0.0003 | 0.0000 | 0.0000 |
| | 12 | 0.0005 | 0.0003 | 0.0000 | 0.0000 |

*Note:* All variables are included in levels except for Wal-Mart and Target that are included in differences. Each cell shows a response of the row variable to a shock in column variable (at a given time).

Finally, a positive shock in the number of Target stores seems to deter Wal-Mart from locating nearby although this response is not statistically significant. On the other hand, Target has a positive response to a positive shock in the number of Wal-Mart stores. This supports the “follow the leader strategy” that Target has exhibited over the years as commented in (Basker, 2007). However, this effect is not statistically significant.

***3.4.4. Variance Decompositions***

The variance decompositions for the different orderings, presented in Table 3.8 and 3.9, are in line with the above results. These tables show the percent variation in the row variable, explained by the column variable. Note, only the total effect accumulated over 10 years is reported, but longer time horizons produced equivalent results. Table 3.8 shows that the "big-box effect" is slightly higher on retail employment than in retail wages as explained by the percent of total variation on these variables (0.131% vs. 0.13%). Table 3.8 also shows that retail employment explains 0.72% of the big-box retailers' total variation 10 periods ahead while retail wage only explains 0.35%. This implies that retail employment is a relatively heavier factor affecting the big-box store location decisions.

Table 3.9 shows that the total variation (10 periods ahead) explained by Wal-Mart's presence on retail wages is higher compared to the variation explained by Target (0.07% vs. 0.0051%). Similarly, Wal-Mart's presence explains 0.2% of the total variation in retail employment, while Target only explains 0.0008%. These results support that Wal-Mart drives the effect in the aggregate "big-box effect" above. In addition, as given by the variance decompositions in Table 3.9, retail employment has a relatively bigger impact on Wal-Mart's location decision compared to retail wages (0.69% vs. 0.64%). On the other hand, retail wage has a higher impact on Target's location decision than retail employment (0.033% vs. 0.0073%).

**Table 3.8. Variance Decompositions ($\boldsymbol{rw_{it}, re_{it}, bx_{it}}$)**

| | *rw* | *re* | $\Delta bx$ |
|---|---|---|---|
| *rw* | 0.931518 | 0.067179 | 0.001303 |
| *re* | 0.007919 | 0.990775 | 0.001306 |
| $\Delta bx$ | 0.003492 | 0.007235 | 0.989273 |

*Note:* Variation in the row variable explained by column variable (10 periods ahead).

**Table 3.9. Variance Decompositions $(rw_{it}, re_{it}, wm_{it}, tg_{it})$**

| | $rw$ | $re$ | $\Delta wm$ | $\Delta tg$ |
|---|---|---|---|---|
| $rw$ | 0.8638 | 0.1355 | 0.0007 | 5.08E-05 |
| $re$ | 0.0196 | 0.9784 | 0.0020 | 7.92E-06 |
| $\Delta wm$ | 0.0064 | 0.0069 | 0.9866 | 8.95E-06 |
| $\Delta tg$ | 0.0003 | 7.33E-05 | 2.62E-06 | 0.9996 |

*Note:* Variation in the row variable explained by column variable (10 periods ahead).

### 3.5. Summary and Concluding Remarks

This chapter complements the earlier work in the literature concerning the big-box retailers', Wal-Mart and Target, impact on retail wages and employment. The previous work findings, albeit conflicting, suggest a significant relationship between these big-box stores' location, retail wage and employment (endogeneity). This chapter attempts to address the endogeneity issue between the big-box retailers' store location decision and their effect on retail wage and employment via the panel VAR approach. The panel VAR modeling approach provides a unifying empirical framework and identification strategy - previously absent in the literature - addressing the big-box stores impact on the local economy.

According to the results, the effect of the big-box stores on retail wage is negative and statistically significant (Figure 3.2). The effect of the big-box retailers on retail employment is positive and significant, although it is relatively smaller in terms of dynamic response, compared to the effect on retail wage rates (Figure 3.2).[6] However, based on the variance decompositions the big-box retailers' presence has a slightly larger impact on retail employment. Regarding the location decision of these big-box stores, results suggest the level of retail employment is a more important factor to consider. The impact of the big-box stores is driven by the effects from Wal-Mart (Figure 3.3). Wal-Mart's individual effect on retail wage and employment is similar to that of

[6] The exact source for these gains in employment in the retail sector is unknown. However, some complementary establishments may yield positive benefits from Wal-Mart's presence , which may have an impact on retail job gains.

the aggregate big-box effect. As shown in Figure 3.3, there exits two opposite impacts of Wal-Mart: (i) the negative effect on retail wages - which is consistent with most of previous literature, and (ii) positive effect on retail employment. The resulting effects of Target stores on retail wage and employment are insignificant.

The big-box retailers' location decision is also of interest. The variance decompositions show that Target's location decision is slightly more heavily affected by retail wage than employment level. Conversely, retail employment has a relatively bigger impact on Wal-Mart's location decision compared to retail wages. As anticipated, Target portrays more of a follower strategy when it comes to location decision for counties that have a Wal-Mart store. Conversely, Wal-Mart may choose to avoid counties with an existing Target store.

# CHAPTER 4

## BIG-BOX RETAILERS AND PERSONAL INCOME GROWTH

### Abstract

This chapter addresses a research gap in the literature by investigating the impact of big-box retailers' presence on personal income growth in U.S. counties between 2000 and 2005, based on neoclassical growth models of cross-country income convergence. Whether big-box retailers have a negative effect on local economic growth has been a permeating question amongst regional developers, policy makers and economists. Wal-Mart's and Target's economic impacts are estimated in regards to the degree in which their individual presence affects personal income growth. Different model specifications are applied in the analysis, including spatial models that control for spatial autocorrelation. Results suggest that counties with the presence of both Wal-Mart and Target stores have experienced slower growth in personal income - even after controlling for spatial autocorrelation. Wal-Mart's individual effect on personal income growth is negative and highly significant. Target's individual effect is also negative, but statistically insignificant after controlling for spatial dependence.

Key Words: Personal Income, Growth, Income Convergence, Spatial Econometrics, Wal-Mart, Target.

JEL Codes: O47, O51, R11

## 4.1. Introduction

As Wal-Mart continues to grow and open more stores nationwide, its impact on local economies remains a concern for the general population, local policy makers, and researchers (Bonanno and Goetz (2012). According to Wal-Mart annual reports[7], Wal-Mart has sustained a positive growth in net sales[8] in spite of recent tumbles in the global economy. In the U.S. alone, Wal-Mart employs over a million associates with a total of 4,516 stores along with more than 600 Sam's Clubs (Wal-Mart, http://corporate.walmart.com/our-story/locations#/united-states). Additionally, similar retail stores such as Target, which is a more upscale big-box retailer, and yet has to some extent mirrored Wal-Mart's growth across the U.S. (Basker, 2007). To date Target has a total of 1,795 stores in the U.S. and 347,000 team members worldwide (Target, 2015).

A large amount of previous research has focused on examining Wal-Mart's effect on local economic outcomes (i.e. employment, wages, poverty level, food prices, etc.) within specific regions, states, counties and localities in the U.S. This is perhaps a byproduct of Wal-Mart's aggressive and large expansion throughout the U.S., its industry leader status, and success over its common competitors such as Kmart and Target (Basker, 2007). In fact, the degree to which Wal-Mart's impact on local economies is quantitatively or qualitatively different from the effect of other "big-box" retailers such as Costco, Target, or Kmart, remains an important open question (Basker, 2007). One exception is Jia (2008), who estimates the effect on small general merchandise stores from both Kmart and Wal-Mart, and concludes that they have similar impacts on the small stores' exit decisions. In this chapter Wal-Mart's and Target's impact is estimated along with the degree to which their individual or aggregate presence affects personal income growth in U.S. counties.

[7] http://stock.walmart.com/investors/financial-information/annual-reports-and-proxies/default.aspx
[8] Net sales is the amount of sales generated by a company after the deduction of returns, allowances for damage or missing goods and any discounts allowed.

Different model specifications are applied in the analysis including a spatial error model to control for spatial autocorrelation.

Some relevant studies in the literature include Keil and Spector (2005), which examines the effect of Wal-Mart on income differentials and unemployment between the black and white populations in Alabama - using county census data from 1980 and 1990. They find that Wal-Mart's presence significantly correlates to lower unemployment for blacks, while the impact on income is trivial after controlling for other socio-economic variables. Basker (2007) explores Wal-Mart's competitive advantage and how its presence affects consumer prices, local labor markets, global and local competitors including Target, suppliers and product selection. Although the study is more of a qualitative analysis and survey of the literature on Wal-Mart, the author emphasizes how Wal-Mart's location decision depends on the local economic conditions. Basker (2011) uses quarterly data for 1997-2006 to estimate the aggregate income elasticity of revenue for Wal-Mart and Target. She finds that during an economic downturn, Wal-Mart's revenues increase whereas Target's revenues decline.

In a related vein, Jantzen et al. (2009) use cointegration techniques and causality tests to examine the relationship between Wal-Mart sales and a set of macro measures of income, employment, production, and prices. They conclude that Wal-Mart's sales soar during periods of slow economic growth and decline during periods of economic boom. However, their study uses aggregate data (national level) and does not account for other competing retailer's economic impact. It is important to note that for the average consumer, Wal-Mart is perceived as a discount haven. As such, Wal-Mart's entry is considered to have an overlapping effect, since its lower prices indirectly influence competing stores to lower their own prices. This indirect effect is accounted to vary from 3% in overall to 13% for specific items (Basker, 2005, Hausman and Leibtag, 2007).

In general, the impact of Wal-Mart's entry on local retailers' sales is considered to be negative for direct competitors although some complementary establishments may reap positive benefits from Wal-Mart's presence (Ailawadi, et al., 2010, Artz and Stone, 2006, Irwin and Clark, 2006, Jia, 2008, Stone, 1988, Stone, 1997, Stone, 1995). Similarly, other studies link Wal-Mart and large discount chains' presence to the closure of small shops in downtown and local main streets, declines in employment and wages, community disruption and higher poverty (Freeman, 2003, Goetz and Swaminathan, 2006, McGee and Gresham, 1996, Quinn, 2005). However, Barnes, et al. (1996) do not find a negative effect on the number of establishment nor their sales due to Wal-Mart presence in Northeast markets.

On the other hand, some researchers have focused specifically on the effect from Wal-Mart's presence on local retail employment and wages (Basker, 2005, Bernstein and Bivens, 2006, Dube, Lester and Eidlin, 2007, Hicks, 2007, Ketchum and Hughes, 1997, Neumark, Zhang and Ciccarella, 2008). While some authors find modest gains in employment as a result of Wal-Mart's entry (Basker, 2007, Basker, 2005, Drewianka and Johnson, 2006, Hicks and Wilburn, 2001, Ketchum and Hughes, 1997), others argue on decreasing employment (Hicks, 2008, Hicks, 2007, Neumark, Zhang and Ciccarella, 2008). Most studies find little to modest increases in retail wages for areas with a Wal-Mart (Goetz and Shrestha, 2009, Hicks, 2008, Hicks and Wilburn, 2001, Ketchum and Hughes, 1997). However, Neumark, Zhang and Ciccarella (2008) find slight decreases on retail payroll (wages) due to Wal-Mart's presence. Chapter 3 in this dissertation finds that the big-box retail stores increases the number of jobs while decreases retail wages.

This chapter intends to address a research gap in the literature by investigating the local economic impact of Wal-Mart's and Target's presence on county level personal income growth in the U.S. counties, within the 48 contiguous states. To the best of the authors' knowledge, there is no study that investigates the effect from Wal-Mart and Target stores in regards to the degree in which their individual or aggregate presence affects personal income growth in U.S. counties - while controlling for spatial autocorrelation. The empirical model is built upon the theoretical framework of neoclassical growth models of cross-county income convergence (Barro and Sala-i-

Martin, 1992, Mankiw, et al., 1992). The research objective is to determine if there is a relationship between personal income growth and the presence of Wal-Mart and Target stores.

## 4.2. Methodology and Data

### *4.2.1. Income Growth Model*

The model in equation (4-1) is based on neoclassical growth models of cross-country income convergence, i.e., poor countries tend to grow faster and catch up with rich countries, as in Barro and Sala-i-Martin (1992), and Mankiw, Romer and Weil (1992), to evaluate the impact of Wal-Mart's and Target's presence on personal income growth:

(4-1) $$\delta_i = \beta_0 + \beta_1 wm_{2000,i} + \beta_2 trgt_{2000,i} + \beta_3 lnY_{2000,i} + \boldsymbol{\theta}' \, \boldsymbol{E}_{2000,i} + \boldsymbol{\varphi}' \, \boldsymbol{X}_{2000,i} + \sigma_s + \epsilon_i,$$

where $\delta_i = ln(Y_{2005,i}) - ln(Y_{2000,i})$ is the personal income growth rate between 2000 and 2005 in county i. The base year is 2000. The period of analysis is selected according to the data availability at the county level for Wal-Mart and Target, as well as to the socio-demographic data from the US Census. The term $wm_{2000,i}$ is the number of Wal-Mart stores in county $i$ in year 2000; $trgt_{2000,i}$ is the number of Target stores in county i in year 2000, and $ln\,Y_{2000,i}$ is the natural log of per capita personal income in year 2000; $\boldsymbol{E}_{2000,i}$ is the vector of shares of earnings for the county industry sectors considered in the analysis; $\mathbf{X}_{2000,\mathrm{i}}$ is a vector of socio-economic and demographic variables measured in year 2000; $\sigma_s$ is the state-specific dummies for the fixed effect and $s = 2, \ldots, 48$; $\epsilon_{\mathrm{i}}$ is the error term. The coefficients $\beta_1$ and $\beta_2$ in model in (4-1) test the statistical significance of the effect of Wal-Mart's and Target's presence on personal income growth.

### *4.2.2. Controlling for Spatial Dependence*

Analysis of regression relationships with sample data that is spatial in nature can produce spurious estimation results. This is because spatial data typically violates the assumption made by ordinary regression models in which each observation is assumed to be independent of other observations (LeSage, 2014). In Anselin and Bera (1998), spatial autocorrelation is loosely defined as the coincidence of value similarity with locational similarity. More formally, the presence of spatial autocorrelation can be illustrated by the following moment condition:

(4-2) $$Cov(y_i, y_j) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0 \quad \text{for } i \neq j,$$

where $y_i$ and $y_j$ are observed random variables for location $i$ and $j$ in space. The subscripts $i$ and $j$ might be points measured as latitude and longitude (e.g., metropolitan areas, stores locations), or areal units (e.g., counties, states).

For the income growth model proposed in (4-1) spatial econometrics is used to account for the presence of spatial effects in the regression analysis. An extensive overview of the relevant methodology is beyond the scope of this chapter, but technical aspects of spatial regression diagnostics are reviewed in Anselin (1988), Anselin and Bera (1998), Anselin (2001), LeSage and Pace (2009), LeSage (2014), among others. To test the presence of spatial dependence in the sample data, the Moran's I test as in Cliff and Ord (1972) is calculated. As discussed in Anselin and Bera (1998), the test was originally developed as a two dimensional analog of the test of significance of the serial correlation coefficient in univariate time series. In Cliff and Ord (1972), Moran's I statistics is formally expressed as:

(4-3) $$I = \frac{N}{S_0}\left(\frac{e' \boldsymbol{W} e}{e' e}\right),$$

where $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ is a vector of least squares residuals, $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}' \boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}' \boldsymbol{y}$, $\boldsymbol{W}$ is the spatial weights matrix based on contiguity or distance, $N$ is the number of observations, $S_0$ is a standardized factor that is equal to the sum of spatial weights, or $\Sigma_i \Sigma_j w_{ij}$. Here $S_0$ simplifies to $N$ for a row-standardized weights matrix $\boldsymbol{W}$, because each row sum equals 1. The Moran's I statistic then becomes

(4-4) $$I = (\frac{e' \boldsymbol{W} e}{e' e})$$

A statistically significant Moran's I statistic suggests a problem with spatial autocorrelation. Different model specifications may be used once spatial autocorrelation is detected in order to address this matter. These include the spatial lag and spatial error regressions as shown in equations (4-5) and (4-7), respectively.

(4-5) $$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is a $N$ by 1 vector of the dependent variable, $\boldsymbol{W}\boldsymbol{y}$ is the spatially lagged dependent variable with weights matrix $\boldsymbol{W}$, $\rho$ is the spatial autoregressive parameter, an $N$ by $K$ matrix of explanatory variables is given by $\boldsymbol{X}$, $\boldsymbol{\beta}$ is a $K$ by 1 vector of coefficients, and $\boldsymbol{\epsilon}$ is a $N$ by 1 vector of errors. The reduced form of the spatial lag model is expressed as:

(4-6) $$(1 - \rho \boldsymbol{W}) \boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $(1 - \rho \boldsymbol{W}) \boldsymbol{y}$ is a spatially filtered dependent variable (i.e., with the effect of spatial autocorrelation removed). This is somewhat analogous to first differencing the dependent variable in time series. However, the $\rho = 1$ scenario is not allowed in the parameter space of (4-6). Correspondingly, the spatial autoregressive parameter $\rho$ must be explicitly estimated. The independent variables explain the variation in the dependent variable that is not explained by the neighbors' value or autoregressive parameter.

As described in Anselin and Bera (1998), a second way to incorporate spatial autocorrelation in a regression model is to specify a spatial process for the disturbance terms. The authors present the most common specification as a spatial autoregressive process in the error terms:

(4-7) $$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

This is a linear regression with error vector $\boldsymbol{\epsilon}$, so that:

(4-8) $$\boldsymbol{\epsilon} = \lambda \boldsymbol{W} \boldsymbol{\epsilon} + \boldsymbol{\xi}$$

In (4-8) $\lambda$ denotes the spatial autoregressive coefficient for the lag of the error $\boldsymbol{W\epsilon}$ (this is a different notation to that of the spatial autoregressive coefficient ρ of a spatial lag model), $\boldsymbol{\xi}$ is an uncorrelated and homoscedastic error term (without loss of generality). Alternatively, (4-7) may be rewritten as

(4-9) $$\boldsymbol{y} = \boldsymbol{X\beta} + (1 - \lambda\boldsymbol{W\epsilon})^{-1}\boldsymbol{\xi}, \quad \text{or}$$

(4-10) $$(1 - \lambda\boldsymbol{W\epsilon})\boldsymbol{y} = (1 - \lambda\boldsymbol{W\epsilon})\boldsymbol{X\beta} + \boldsymbol{\xi}$$

Despite the power in Moran's I statistic to detect model misspecifications (and not simply spatial autocorrelation), it is not suitable in suggesting the alternative model specification that should be used. As such, the spatial regression model selection is done using Lagrange Multiplier test statistics. Although initially the range of available test statistics for spatial autocorrelation may be puzzling, one can follow a fairly intuitive process or decision rule for a spatial regression model selection. This process is summarized in Figure 4.1. For the spatial model specification selected, controls identical to those in (4-1) were included.

**4.3. Data**

The selection of variables for the model in (4-1) follows the economic growth literature including Acemoglu, Johnson and Robinson (2003), Bloom, Canning and Malaney (2000), Dixit (1973), Higgins, Levy and Young (2006), Lucas (1988), Malmberg (1994), Mankiw, Romer and Weil (1992), Quigley (1998), Zak and Knack (2001), James and Aadland (2011). The period 2000-2005 is selected based on data availability at the county level for Wal-Mart and Target stores, along with the socio-demographic data from the US Census. A control variable with the share of earnings in the high-tech industry sector is introduced for 2000, in order to control for the dot-com bubble.

The sample data covers 3,050 counties in the 48 contiguous States of the U.S., after dropping 94 counties due to missing data. Personal income and population data are obtained from the Bureau of Economic Analysis (BEA). The BEA defines personal income as the income received by persons from all available sources. It is the sum of net earnings by place of residence, property income, and personal current transfer receipts.

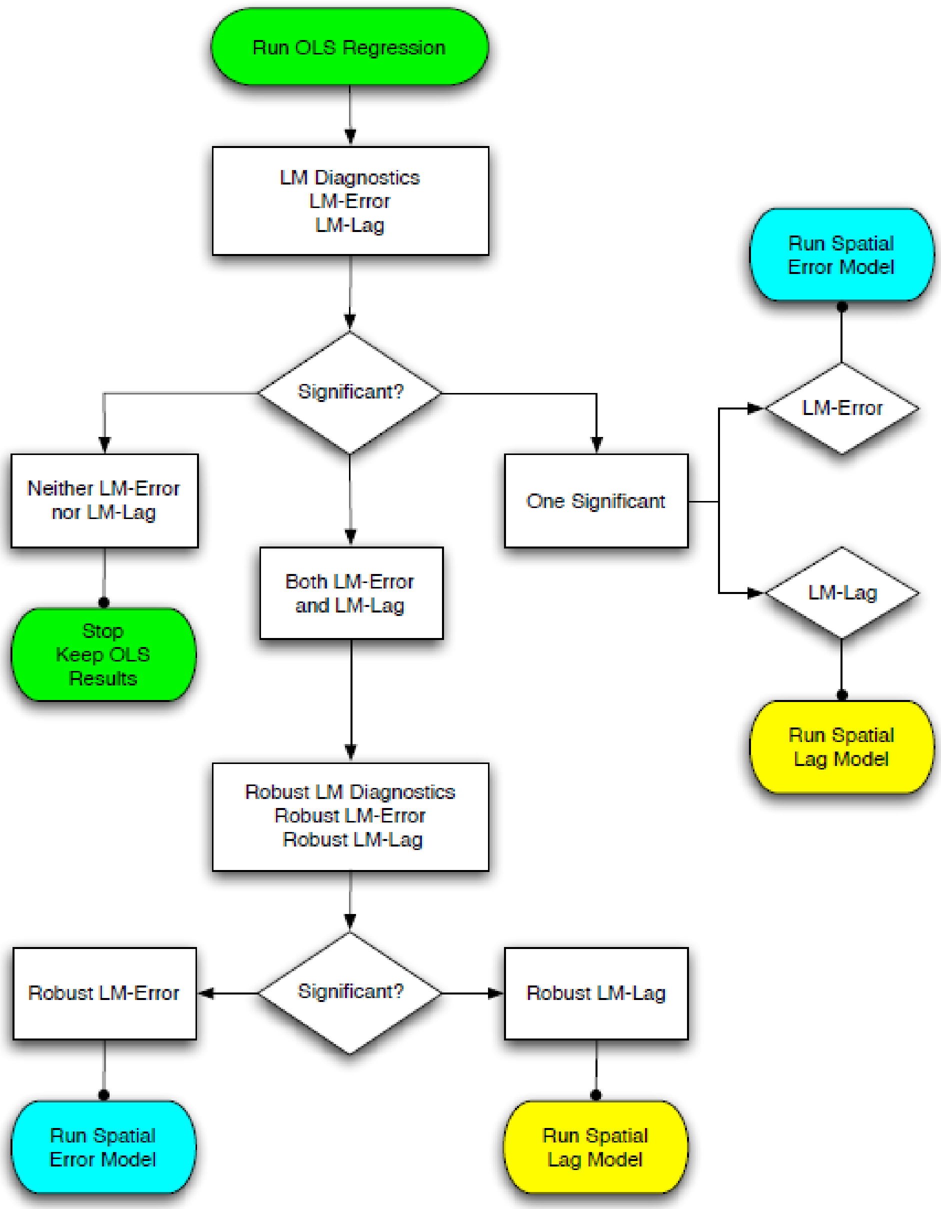


**Figure 4.1. Spatial regression decision process**
*Source:* Exploring Spatial Data with GeoDa: A Workbook,
https://geodacenter.asu.edu/system/files/geodaworkbook.pdf, p. 199

For the sample, personal income (per capita) in current dollars is deflated using the 2009 GDP deflator. Industry earnings[9] (percentage of total industry earnings) are obtained from the US Census Bureau's County Business Patterns database for the natural resource sectors (the sum of agriculture, forestry, fishing, and mining) and high-tech sectors.[10] Other socio-economic variables such as percentage of population with only high school diploma, percentage of population with a college degree or higher, poverty rate, and population density (metro dummy)[11] are compiled from the U.S. Census Bureau database. Similarly, longitude and latitude data for the spatial analysis are compiled from the U.S. Census Bureau.

The Wal-Mart variable measures the number of stores in year 2000 at the county level. The number of Wal-Mart stores during this period is compiled from Holmes' (2011) database which is available at http://www.econ.umn.edu/~holmes/data/WalMart/, and normalized by population (per 100,000 inhabitants). Wal-Mart made a public file which lists a Wal-Mart store, address, store number, store type (supercenter or regular one), and opening data in November 2005. Holmes (2011) combines these data with additional information posted at Wal-Mart website and lists the opening date for each store. For the analysis in this chapter, the store count by county in year 2000 is generated using Holmes' data set.

Target store count was generated in a similar fashion using the Target Store Openings data available at FLOWINGDATA https://flowingdata.com/2009/10/22/target-store-openings-since-the-first-in-1962-data-now-available. A depiction of Wal-Mart and Target store count is presented in Figure 4.2.

---

[9] Earnings refer to payroll data as defined by the US Census County Business Patterns.

[10] The NAICS code considered in formulating the share of earnings for the "high-tech" most relevant industries to the dot-com bubble include 334 (Computer and Electronic Product Manufacturing), 51 (Information), 5415 (Computer Systems Design and Related Services), 5417 (Scientific Research and Development Services), 5232 (Securities and Commodity Exchanges), 8112 (Electronic and Precision Equipment Repair and Maintenance).

[11] Metro dummy= 1 if population per square mile in 2000 exceeds 300, else zero following James and Aadland (2011).

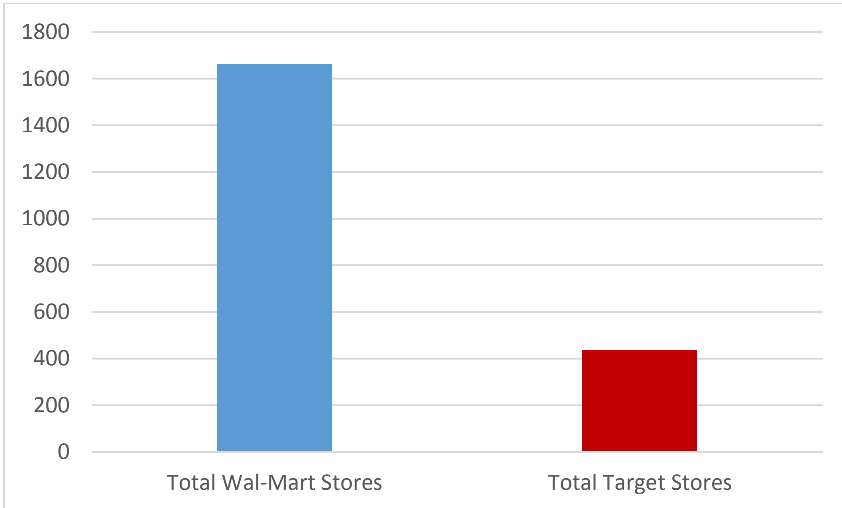


**Figure 4.2. Wal-Mart and Target store count in year 2000**
*Source*: Figure generated with author's sample data

Definition of the variables in the model and descriptive statistics are presented in Table 4.1. The average income growth rate is 7% (median 6%) between 2000 and 2005 across the counties in the sample. The average number of Wal-Marts per 100,000 inhabitants is 1.51 in 2000, while for Target this figure is only 0.16. The average share of earnings in resource sector is about 2.2%, while the high-tech sector share of earnings accounts for 3% across counties. In 2000, an average of 16% of the population held at least a college degree, while the average poverty rate sat at 14%.

## 4.4. Empirical Results

### *4.4.1. Least Squares Estimation*

The robust standard error OLS results from equation (4-1) are shown in Table 4.2. As shown in Table 4.2, five different regression models are estimated to control for initial income, shares of industry earnings (resource, high-tech), human capital, age structure, ethnicity, poverty, and population density (metro dummy). In all five regressions, state-specific fixed effects were included, but estimated coefficients are not reported to save space. Instead, F statistics for joint significance are reported in.

**Table 4.1. Definition of Variables and Summary Statistics**

| Variable | Definition | Mean | Std. Dev. | Min | Med. | Max |
|---|---|---|---|---|---|---|
| $\delta$ | Growth in personal income (2000-2005) | 0.07 | 0.08 | -0.31 | 0.06 | 0.71 |
| $ln(Y_{2000})$ | Personal income per capita in 2000 | 10.25 | 0.22 | 9.42 | 10.24 | 11.52 |
| Wal-Mart | Number of Wal-Mart Stores (per 100,000 people) in 2000 | 1.51 | 2.07 | 0 | 0.64 | 16.50 |
| Target | Number of target Stores (per 100,000 people) in 2000 | 0.16 | 0.83 | 0 | 0.00 | 24.70 |
| Resources | Percent of earnings in agriculture, forestry, fishing, mining in 2000 | 0.02 | 0.07 | 0 | 0.00 | 0.96 |
| High-tech | Percent of earnings in high-tech industries in 2000 | 0.03 | 0.04 | 0 | 0.01 | 0.40 |
| High School | Percent of population with only high school education in 2000 | 0.35 | 0.07 | 0.11 | 0.35 | 0.53 |
| College | Percent of population with at least a college degree in 2000 | 0.16 | 0.08 | 0.05 | 0.14 | 0.61 |
| Young | Percent of population that is less than 18 years old in 2000 | 0.26 | 0.03 | 0.15 | 0.25 | 0.45 |
| Old | Percent of population that is at least 65 years old in 2000 | 0.15 | 0.04 | 0.02 | 0.14 | 0.35 |
| Poverty | Percent of population at or below poverty line in 2000 | 0.14 | 0.07 | 0 | 0.13 | 0.57 |
| Ethnicity | Percent of Caucasian population in 2000 | 0.82 | 0.19 | 0.02 | 0.90 | 1.00 |
| Metro | Dummy = 1 if population per square mile in 2000 exceeds 300, else zero | 0.10 | 0.30 | 0 | 0.00 | 1 |

*Notes:* $N = 3050$ observations for all variables in the sample. Figures have been rounded to the nearest decimal.

The coefficient for the Wal-Mart variable is negative and significant in all models, and it suggests that counties with a Wal-Mart presence have grown slower in terms of personal income. This negative relationship between Wal-Mart and personal income growth can also imply that more Wal-Mart stores might slow down the local economic growth due to the possible closure of small downtown and main street stores, leading to declines in employment and wages, as noted in McGee and Gresham (1996), Freeman (2003), Quinn (2005) and Goetz and Swaminathan (2006).

**Table 4.2. Robust Estimates for Income Growth**

| Variable | Model 1 Coeff. (std. err.) | Model 2 Coeff. (std. err.) | Model 3 Coeff. (std. err.) | Model 4 Coeff. (std. err.) | Model 5 Coeff. (std. err.) |
|---|---|---|---|---|---|
| Constant | 0.0965*** | 1.3798*** | 1.3506*** | 1.8983*** | 1.8704*** |
| | (0.0060) | (0.0862) | (0.0878) | (0.1914) | (0.1900) |
| Wal-Mart | -0.0018*** | -0.0024*** | -0.0023*** | -0.0019*** | -0.0018*** |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| Target | -0 .0054*** | -0.0025* | -0.0026* | -0.0032* | -0.0031* |
| | (0.0019) | ( 0.0014) | (0.0014) | (0.0016) | (0.0016) |
| $\ln(\mathrm{Y}_{2000})$ | | -0.1264*** | -0.1237*** | -0.1930*** | -0.1902*** |
| | | (0.0084) | (0.0086) | (0.0195) | (0.0194) |
| Resources | | | 0.0659** | 0.0856*** | 0.0843*** |
| | | | (0.0263) | (0.0272) | (0.0273) |
| High School | | | | -0.0502 | -0.0511 |
| | | | | (0.0452) | (0.0452) |
| College | | | | 0.3353*** | 0.3595*** |
| | | | | (0.0489) | (0.0495) |
| Young | | | | 0.3150*** | 0.3151*** |
| | | | | (0.0733) | (0.0729) |
| Old | | | | 0.3552*** | 0.3475*** |
| | | | | (0.0548) | (0.0546) |
| Poverty | | | | 0.0023 | 0.00451 |
| | | | | (0.0454) | (0.0454) |
| Ethnicity | | | | -0.0069 | -0.0075 |
| | | | | (0.0130) | (0.0130) |
| Metro | | | | -0.0074 | -0.0036 |
| | | | | (0.0045) | (0.0045) |
| High-tech | | | | | -0.1306*** |
| | | | | | (0.0385) |
| $N = 3050$ | | | | | |
| F stat. state FEs | 25.79*** | 24.16*** | 23.45*** | 13.51*** | 13.90*** |
| $\mathrm{R}^2$ | 0.230 | 0.317 | 0.319 | 0.355 | 0.357 |

*Notes:* Superscripts ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are in parentheses. The state fixed-effects estimates are not shown. However, the F statistic are reported for the joint significance of the state fixed-effects. The $R^2$ values are reported for each OLS estimation.

The coefficient for the Target store variable also shows a negative relationship with respect to personal income growth. One can interpret these results as Target also having a negative impact on local personal income growth. However, the target coefficient is only significant at the 10%

significance level for most of the models shown in Table 4.2. Meanwhile, other estimates for equation (4-1) show that the logged initial income has a negative and statistically significant coefficient. This is consistent with the theory of the conditional income convergence (Higgins, Levy and Young, 2006), i.e., poor regions grow faster.

The share resource earning coefficient is positive. This is contrary to the “curse of natural resources” argued in James and Aadland (2011). This curse refers to the link found in the resource literature between lower economic growth and natural resource dependence. The coefficient on high-tech earnings, as expected, is negative and statistically significant. This indicates that counties with a larger share of earnings in high-tech industries have experienced slower income growth. This is likely as a result of the dot-com bubble burst during early 2000s. Human capital variables such as college (percent of population with at least a college degree in 2000) have a positive and significant influence on personal income growth as suggested in Higgins, Levy and Young (2006), Lucas (1988), and Mankiw, Romer and Weil (1992). The poverty rate and the ethnicity variables have the expected sign. The density (metro dummy) coefficient is negative although not significant.

### *4.4.2. Spatial Models*

Following the spatial regression decision process outlined in Figure 4.1, the OLS regression model is estimated along with the diagnostics for spatial dependence. The OLS model specification follows equation (4-1). The spatial weight matrix used in the spatial analysis is a distanced-based spatial weight matrix, with a distance band of 90.84 miles. This is the minimum distance threshold ensuring that each county will have at least one neighbor. The county centroids are approximated using GeoDa, since the longitude and latitude data is unprojected.

The regression diagnostics reveal considerable non-normality and heteroscedasticity. This indicates the presence of heteroskedastic errors, possibly as a result of spatial autocorrelation. The diagnostics for spatial dependence are given in Table 4.3. A total of five test statistics are reported.

**Table 4.3. Diagnostics for Spatial Dependence**

| County Distance Weight Matrix (row-standardized) | | |
|---|---|---|
| Test | Statistic Value | P-value |
| Moran's I (error) | 19.50 | 0.000 |
| Lagrange Multiplier (lag) | 279.26 | 0.000 |
| Robust LM (lag) | 55.96 | 0.000 |
| Lagrange Multiplier (error) | 231.94 | 0.000 |
| Robust LM (error) | 8.63 | 0.003 |

*Notes:* The distanced band used in the weight matrix is 90.84. This is the minimum threshold distance ensuring that each county will have at least one neighbor.

The first statistic is Moran's I. It is highly significant. As discussed in section 4.2.2 and indicated in Figure 4.1, this suggests a problem with spatial autocorrelation. Lagrange Multiplier statistics are used to determine which spatial model specification should be utilized. In the diagnostic output, four Lagrange Multiplier test statistics are presented. The first two (LM-Lag and Robust LM-Lag) regard the spatial lag model as the alternative. The following two (LM-Error and Robust LM-Error) refer to the spatial error model as the alternative.

As in the decision process outlined in Figure 4.1., the two standard (i.e., not the robust forms) LM-Lag and LM-Error test statistics are considered first. Since both statistics reject the null, the robust forms of the test statistics are considered next. Because both robust statistics are highly significant, the spatial model specification is selected under the basis of the largest value for the test statistic, as suggested in Anselin (2004). Accordingly, the spatial lag specification is selected.

The spatial lag model is estimated by maximum likelihood methods. The model follows a similar structure as in (4-6). The estimates and measures of fit are also given in Table 4.4. The pseudo-$R^2$ is not directly comparable with the measure given in the OLS estimation results in Table 4.4. Nonetheless, more appropriate measures of fit are reported (e.g., Log-Likelihood, AIC, and SC). For comparison purposes, the spatial error model is also estimated and reported in Table 4.4.

**Table 4.4. Spatial Analysis for Income Growth**

| Variable | OLS (Model 5) Coeff. (std. err.) | Spatial-lag Coeff. (std. err.) | Spatial-error Coeff. (std. err.) |
|---|---|---|---|
| Constant | 1.8718*** | 1.6870*** | 1.7899*** |
| | (0.1241) | (0.1193) | (0.1230) |
| Wal-Mart | -0.0018*** | -0 .0017*** | -0.0017*** |
| | (0.0006) | (0.0006) | (0.0006) |
| Target | -0 .0027* | -0.0022 | -0.0018 |
| | (0.0016) | (0.0015) | (0.0015) |
| $\ln(Y_{2000})$ | -0.1904*** | -0.1745*** | -0.1828*** |
| | (0.0119) | (0.0115) | (0.0118) |
| Resources | 0.0843*** | 0.0503*** | 0.0325* |
| | (0.0195) | (0.0187) | (0.0195) |
| High School | -0.0495 | -0.0141 | 0.0019 |
| | (0.0426) | (0.0407) | (0.0443) |
| College | 0.3591*** | 0.3289*** | 0.3480*** |
| | (0.0387) | (0.0370) | (0.0390) |
| Young | 0.3155*** | 0.2562*** | 0.2689*** |
| | (0.0594) | (0.0568) | (0.0584) |
| Old | 0.3497*** | 0.2549*** | 0.2599*** |
| | (0.0468) | (0.0450) | (0.0476) |
| Poverty | 0.0049 | -0.0098 | 0.0152 |
| | (0.0384) | (0.0367) | (0.0390) |
| Ethnicity | -0.0075 | 0.0007 | 0.0075 |
| | (0.0125) | (0.0119) | (0.0136) |
| Metro | -0.0039 | 0.0005 | -0.0009 |
| | (0.0052) | (0.0050) | (0.0050) |
| High-tech | -0.1321*** | -0.1097*** | -0.0911** |
| | (0.0389) | (0.0371) | (0.0372) |
| Rho | | 0.4941*** | |
| | | (0.0348) | |
| Lambda | | | 0.5894*** |
| | | | (0.0361) |
| *N* = | 3050 | 3050 | 3050 |
| R-squared | 0.357 | 0.402 | 0.403 |
| Jarque-Bera | 8727.9333 | | |
| P-value | 0.0000 | | |
| Bresch-Pagan Test | 1111.819 | 1024.301 | 1012.833 |
| P-value | 0.000 | 0.000 | 0.000 |
| Log-likelihood | 4058.65 | 4149.60 | 4140.8699 |
| Akaike info criterion | -7997.31 | -8177.20 | -8161.74 |
| Schwarz criterion | -7635.93 | -7809.80 | -7800.37 |
| Likelihood Ratio test | | 181.893 | 164.432 |
| P-value | | 0.000 | 0.000 |

*Note:* Superscripts ***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively. Standard errors are in parentheses. The state fixed-effects estimates are not shown. A row-standardized distance-based weight matrix is used to fit the spatial lag and spatial error model. The distanced band used in the weight matrix is 90.84. This is the minimum threshold distance ensuring that each county will have at least one neighbor.

Comparing these values to those for OLS, one can notice an increase in the value of log-likelihood. Additionally, considering the fit with respect to the added spatially lagged dependent variable, both the AIC and SC decrease relative to OLS estimates. This again suggests an improvement of fit for the spatial lag specification over least squares.

The spatial autoregressive coefficient (rho) is 0.4941, and highly significant. Similar to the OLS results in Table 4.4, for the spatial lag model the coefficient for Target is negative but not significant (at the 5% level). This means that Target's presence alone may not have an impact on personal income growth after controlling for spatial dependence. The coefficient on Wal-Mart for the spatial lag model although slightly smaller relative to the OLS results, it is also negative and highly significant. This implies that Wal-Mart presence in year 2000 had a negative impact on personal income growth between 2000 and 2005.

All the other coefficients are similar (albeit smaller in absolute value) to the OLS; except for poverty rate, ethnicity and metro dummy that changed signs, and high school variable that becomes statistically insignificant. Overall, the explanatory power of the model in (4-1) that had been attributed to their own in-county value has been improved due to the consideration of neighboring counties. The coefficient of the spatially lagged dependent variable picks up this effect.

A limited number of diagnostics are provided with the maximum likelihood spatial lag estimation. As shown in Table 4.4, the Breusch-Pagan test for heteroscedasticity is significant suggesting that heteroscedasticity may still be a problem. The likelihood ratio test (Anselin, 2004), as one of the three classic specification tests, contrasts the null model (the least square specification) to the alternative model (spatial lag specification). The resulting LR value of 181.89 indicates that the spatial autoregressive coefficient is significant.

Although the three classic tests are asymptotically equivalent, in finite samples they should follow the ordering: $W > LR > LM$ (Anselin, 2004). For the lag model, the Wald test is $W = 14.22.^2 = 202.21$ (the square of the z-value of the asymptotic t-test (not shown), $LR = 181.89$

(Table 4.4) but LM-lag = 279.26 (Table 4.3). This does not align with the expected order and implies a less than satisfactory model specification so far.

The spatial error model is also estimated to compare the results between the spatial errors and lag model specification (Table 4.4). In terms of coefficient magnitude, sign and significance, the results are analogous to those of the spatial lag model. As emphasized in Anselin (2004), this highlights the difficulties in discriminating between the two spatial models. Nonetheless, the results confirm the direction taken as per the decision rule given in Figure 4.1. For example, the value of the log likelihood in the spatial lag model (4149.6) is marginally better than the spatial error model (4140.87). By the same token, the AIC is lower for the spatial lag model (-8177.2) compared to the error model (-8161.74). Nevertheless, the close similarity between the two models' results and the indication of remaining specification problems advocates further refinement of the model specification.

**4.5. Summary and Concluding Remarks**

Whether the big-box retailers' presence, particularly Wal-Mart and Target, have a negative impact on local economic growth has been a permeating question amongst regional developers, policy makers and researchers. This chapter examines the relationship between the presence of these big-box stores and personal income growth at the county level between 2000 and 2005. Wal-Mart and Target stores' impacts are estimated along with the degree to which their individual presence affects personal income growth at the level of U.S. counties. Different model specifications are applied in the analyses, including a spatial model to control for spatial autocorrelation. Empirical results suggest that counties with Wal-Mart and Target stores have experienced slower growth in personal income. After controlling for spatial autocorrelation, Wal-Mart seems to drive the negative impact. The impact of Target is also negative, but insignificant.

Even though the spatial model improves the fit of the model, further diagnostics on the spatial model specification (Table 4.4) indicate some possible remaining misspecifications issues.

Presumably, possible endogeneity between the big-box retailers' location decision and personal income growth may be a source of misspecification. The next chapter addresses this issue by investigating the dynamic relationship between the big-box retailers and personal income growth using the panel VAR approach.

# CHAPTER 5

## BIG-BOX RETAILERS AND INCOME GROWTH: A PANEL VAR ANALYSIS


### Abstract

This chapter attempts to investigate the dynamic interaction between big-box retailers, Wal-Mart and Target, and personal income growth over time at the county level for the period of 1987-2005 - using the panel Vector Autoregression (VAR) approach. Because natural resource endowment and the structure of the economy are important to income growth, the earning shares of natural resources and manufacturing sectors are included in the analysis, assuming that all the variables in the system are endogenous to one another. The results suggest that these big-box retailers negatively affect personal income growth over time, while the personal income growth has no significant effect on the number of big-box retailers in a particular county.


Key Words: Personal Income, Growth, Panel VAR, Big-Box Retailers, Wal-Mart, Target.

JEL Codes: O47, O51, R11

### 5.1. Introduction

The sizable growth across the U.S. over the past two decades of Wal-Mart and other big-box retailers like Target, continues to be a concern for the general public, local policy makers and researchers (Basker, 2007, Bonanno and Goetz, 2012). The past literature has yet to properly analyze the dynamics of the economic relationships between the growing presence of Wal-Mart and Target stores and their effect on the local economy, specifically personal income and its growth. Jantzen, Pescatrice and Braunstein (2009) examine the relationship between Wal-Mart's sales in the U.S. and a set of macro variables such as income (GDP), employment, production and prices. Although their analysis looks into the dynamics of the variables of interest through cointegration techniques and causality tests, the level of data used is quite aggregated, at a national level, and the presence of other big-box retailers are not accounted for.

Bonanno and Goetz (2012) survey the recent literature on Wal-Mart and note the need for a unifying empirical framework and identification strategy to deal with the endogeneity issue of the company's store location decision and the local economy. As discussed in the previous chapters, potential endogeneity between the big-box retailers' location decision and local economic outcomes i.e., personal income growth, may be a source of misspecification when examining the effect of these retailers on the local economy. In this chapter, the dynamic inter-relationships between big-box retailers and personal income growth is investigated at the county level for the period 1987-2005. Given that natural resource endowment and the structure of the economy are important to income growth, the earning shares from natural resources sectors and manufacturing sectors are likewise included in the analysis.

The analysis is done using a panel vector autoregressive (panel VAR) approach, whereby the dynamic interrelationship among variables of interest is examined assuming that all the variables in the system are endogenous to one another. Hence, via a reduced-form VAR model, estimation results do not require strong assumptions that are necessary in models that use questionable instruments to deal with endogeneity. Additionally, the analysis of impulse response functions (IRFs) and variance decompositions allows us to separate the response of personal income growth, and big-box retailers, i.e., Wal-Mart and Target, to shocks from each of the variables of interest.

The results suggest that the big-box retailers negatively affect personal income growth over time, while personal income growth has no significant effect on the number of the big-box retailers. Moreover, the results indicate that the presence of these big-box stores have a small but significant positive effect on the county share of resource earnings. Also, contrary to popular belief, the big-box stores have a positive impact on the share of manufacturing earnings, too. However, the shares of resource and manufacturing earnings have no significant impact on the big-box stores' location decision.

**5.2. Data**

County level data for the 48 contiguous states stretching over the period of 1987-2005 is compiled for the analysis. The sample contains data on personal income, resource and manufacturing earnings, and the number of Wal-Mart and Target stores. Data observations were compiled up to 2005, based on county level data availability for all variables of interest. Panel data endows the researcher with information on the cross sectional and time series dimension. The resulting sample is a panel with small time variability (T) and large cross-sectional variability (N).

The usage of panel data affords the researcher with an increase in the number of observations (and degrees of freedom) and a reduction in possible collinearity amongst the explanatory variables (Hurlin and Venet, 2003). Since many policies related to economic growth and development are formulated at the county level, this study uses counties as the cross-sectional units of analysis (Carlino and Mills, 1987, Deller, et al., 2001, Rupasingha, Goetz and Freshwater, 2002). A total of 3,047 counties are considered in the analysis with some counties dropped due to missing information.

Personal income and population data are obtained from the Bureau of Economic Analysis (BEA). As defined by the BEA, personal income is the income received by persons from all available sources. It is the sum of net earnings by place of residence, property income, and personal current transfer receipts. Before calculating the annual income growth variable, personal income (per capita) in current dollars is deflated by using the 2009 GDP deflator. Industry earnings data are obtained from the US Census Bureau's County Business Patterns database for the resource or primary sector (the sum of agriculture, forestry, fishing, and mining) and the secondary or manufacturing sector.

In the sample, the researcher calculates the county's resource and manufacturing earnings[12] share as a percentage of total industry earnings, respectively. The industry earnings (percentage of total industry earnings) are obtained from the US Census Bureau's for the natural resource sectors (the sum of agriculture, forestry, fishing, and mining) and manufacturing. As explained in the previous chapters, the number of Wal-Mart stores during this period is compiled from Holmes (2011) database, which is available at http://www.econ.umn.edu/~holmes/data/WalMart/, and normalized by population (per 100,000 inhabitants). For the analysis store count by county and year is generated using Holmes data set.

Target store count was generated in a similar manner using the Target Store Openings data available at FLOWINGDATA (https://flowingdata.com/2009/10/22/target-store-openings-since-the-first-in-1962-data-now-available). A variable adding together the annual store count for Wal-Mart and Target is developed for each county. This is the big-box variable, which is normalized by

[12] Earnings refer to payroll data as defined by the US Census County Business Patterns.

population (per 100,000 inhabitants). This gives the aggregate number of big box stores (including Wal-Mart and Target) per 100,000 people in the county.

Summary statistics and variable definitions are shown in Table 5.1. For the entire sample, the average personal income is $26,153 while the average annual growth is 2%. The average number of big-box stores is close to 1 (0.90). When normalized by county population in 100,000, the average number of big-box stores is 1.51. On the other hand, the average county share of earnings for resource and manufacturing is 3% and 25%, respectively.

**Table 5.1. Summary Statistics**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Deflated personal income (dollars) | 26,153 | 6,752 | 6,724 | 105,132 |
| Deflated personal income growth | 0.02 | 0.05 | -0.85 | 0.85 |
| Big box stores | 0.90 | 1.94 | 0 | 68 |
| Big box stores per 100,000 people | 1.51 | 2.25 | 0 | 43.22 |
| Share of manufacturing earnings | 0.25 | 0.19 | 0 | 0.93 |
| Share of resource earnings | 0.03 | 0.07 | 0 | 0.98 |
| County population (persons) | 85,878 | 282,128 | 411 | 9,793,263 |
| *N = 57893* | *T = 19* | | | |

*Note:* The panel contains 3047 counties.

### 5.3. Methodology: Panel VAR

The previous chapter has discussed how local economic conditions, i.e., personal income growth, may be affected by the big-box retailers' location decisions. The underlying endogeneity among the variables of interest- personal income growth, resource and manufacturing earnings share, and the number of big-box retailers, may lead to misspecification issues. The panel VAR methodology fits the purpose of this chapter because panel VAR requires no a priori assumptions regarding the relationship among the variables.

The panel VAR model follows the works of Holtz-Eakin, Newey, and Rosen (1988) and Love and Zicchino (2006). The panel VAR technique combines the traditional VAR approach (in which all variables within the system are considered endogenous), and the panel structure, which allows for unobserved individual heterogeneity (Love and Ariss, 2014, Love and Zicchino, 2006). While the panel VAR framework permits endogenous relationships amongst the variable that enter a system of equations, the short-run dynamic relationships may likewise be later identified. (Koutsomanoli-Filippaki and Mamatzakis, 2009).

The dynamics of the relationship can be explained graphically using impulse response functions (IRF). IRF describes variable's response to another variable's innovation (shock) within the system while holding all others constant (Hamilton, 1994). The focus is on the orthogonalized IRF. These show the reaction of one variable of interest to a shock in another variable of interest. Therefore, one can isolate, for example, the response of personal income growth to a random shock in big-box stores. Variance decompositions are also reported, which show the percent variation in one variable that is explained by the shock or innovation to another variable accumulated over time (Hamilton, 1994). In other words, the variance decompositions give the magnitude of the total effect.

The panel VAR can take the following reduced form:

(5-1) $$\boldsymbol{X}_{it} = \boldsymbol{\Gamma}(L)\boldsymbol{X}_{it} + \boldsymbol{\Phi}_i + \boldsymbol{\zeta}_t + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{X}_{it} = [x_{it}^1, x_{it}^2, \dots, x_{it}^p]$ is a vector containing the variables of interest. The $i$ subscript stands for county and the $\mathrm{t}$ subscript denotes the time period. $\boldsymbol{\Gamma}(L)$ is a polynomial matrix in the lag operator, that is, $\boldsymbol{\Gamma}(L) = \boldsymbol{\Gamma}_1 L^1 + \boldsymbol{\Gamma}_2 L^2 + \cdots + \boldsymbol{\Gamma}_k L^k$. The model also includes a time-invariant and region-specific element $\boldsymbol{\Phi}_i$. It controls for county-specific effects, which can be unobserved or omitted heterogeneity (e.g., climate, geographical location, land-use policy, etc.). The time-specific element is given by $\boldsymbol{\zeta}_t$. It controls for potential shocks that are common across counties but may vary over time (e.g., fiscal policies, business cycle effects, technological progress, etc.). Lastly, $\boldsymbol{\varepsilon}_{it} \sim i.i.d.\ N(\mathbf{0}, \Omega)$ is a vector of idiosyncratic errors.

The VAR approach with panel data requires the same underlying structure for each cross-sectional unit. In practice, however, such constraint is regularly not met. Therefore, individual heterogeneity in the variables' levels is allowed to overcome this restriction on parameters. This is given by the model's fixed effects $\boldsymbol{\Phi}_i$. Nevertheless, the estimator for the fixed effects is not consistent in a dynamic panel. Due to lags of the dependent variables, the fixed effects are correlated with the regressors (Nickell, 1981). Hence, using the mean-differencing approach to get rid of the fixed effects would produce biased coefficients.

To eliminate the fixed effects, the forward mean-differencing approach (i.e., the Helmert procedure) is used as in Arellano and Bover (1995) and Love and Zicchino (2006). Through this Helmert procedure the forward mean (for each county-year the mean of all available future observations) is removed. The time-specific element $\boldsymbol{\zeta}_t$ is also eliminated during this procedure. Nonetheless, the orthogonality between lagged regressors and the transformed variables is conserved. As a result, one may use the lagged regressors as instruments and equation (5-1) may be estimated through system generalized method of moment (GMM) as in Arellano and Bover (1995). Given that the model in equation (5-1) is just identified (i.e., the number of instruments equals the number of regressors) equation (5-1) can also be estimated using 2 stage least squares (2SLS).

### *5.3.1. The Helmert Transformation*

Let $x_{it}^p$ denote a variable in the vector $\boldsymbol{X}_{it}$. And

(5-2) $$\bar{x}_{it}^p = \frac{1}{T}\sum_{t=1}^{T} x_{it}^p,$$

where $\bar{x}_{it}^p$ is the mean of $x_{it}^p$ over time for the *ith* county. Then,

(5-3) $$\tilde{x}_{it}^p = \frac{1}{T-t}\sum_{n=t+1}^{T} x_{in}^p,$$

where $\tilde{x}_{it}^{p}$ denotes the means calculated from the future values of $x_{it}^{p}$; and $T$ denotes the last period of data for a given series. Let

(5-4) $$\tilde{\epsilon}_{it}^{p} = \frac{1}{T-t}\sum_{n=t+1}^{T} \epsilon_{it}^{p},$$

where $\tilde{\epsilon}_{it}^{p}$ is a similar transformation for $\epsilon_{it}^{p}$; and $\epsilon_{it} = \left[\epsilon_{it}^{1}, \epsilon_{it}^{2}, \dots, \epsilon_{it}^{p}\right]'$. Consequently, the Helmert-transformed versions of $x_{it}^{p}$ and $\epsilon_{it}^{p}$ are expressed as:

(5-5) $$\hat{x}_{it}^{p} = \psi_{it}(x_{it}^{p} - \tilde{x}_{it}^{p}),$$

and

(5-6) $$\hat{\epsilon}_{it}^{p} = \psi_{it}(\epsilon_{it}^{p} - \epsilon_{it}^{p}),$$

where $\psi_{it} = \sqrt{(T-t)/(T-t+1)}$.

In equations (5-5) and (5-6), the Helmert observation for time $t$ is given by difference between the observation at time t and the observations at time $t+1$ through $T$. This implies that the Helmert transformation may not be computed for the last year of data available.

### *5.3.2. Empirical Model*

After the Helmert procedure, the transformed model can be expressed as:

(5-7) $$\widehat{\boldsymbol{X}}_{it} = \boldsymbol{\Gamma}(L)\widehat{\boldsymbol{X}}_{it} + \hat{\boldsymbol{\epsilon}}_{it},$$

where $\widehat{\boldsymbol{X}}_{it} = \left(\hat{x}_{it}^{1}, \hat{x}_{it}^{2}, \dots, \hat{x}_{it}^{p}\right)'$and $\hat{\boldsymbol{\epsilon}}_{it} = \left(\hat{\epsilon}_{it}^{1}, \hat{\epsilon}_{it}^{2}, \dots, \hat{\epsilon}_{it}^{p}\right)'$.

This transformed model expresses all observations as deviations from average future observations. In other words, much larger weight is given to observation that are closer to the start of the time series. In addition, serial correlation is not induced during the Helmert transformation. Therefore, similar properties (i.e., homoscedasticity) in the errors should hold afterwards (Arellano and Bover, 1995). As discussed above, after the transformation, lagged values of the regressors can be used as instrument and system GMM to estimate the coefficients.

Impulse response functions (IRFs) are generated using the residuals from the estimation of all the coefficients in (5-7). As explained above, the IRF can show how an endogenous variable responds to a shock in another variable in the system while holding all other innovations as zero. I use Monte-Carlo simulations to compute the IRF confidence intervals as in Love and Zicchino (2006). That is, in equation (5-7) the coefficients, their corresponding variance-covariance matrix and IRF are randomly drawn. This process is repeated 1000 times and a distribution with its 5$^{th}$ and 95$^{th}$ percentiles is built.

The variance-covariance matrix of the errors is unlikely to be diagonal. Therefore, the residuals are decomposed so that they become orthogonal. This then allows the separation of the shocks to the system's variables. A specific variable ordering is chosen and the Cholesky decomposition is utilized to compute the IRF. In the Cholesky decomposition the series that enter first in the ordering are assumed to have a contemporaneous effect on the subsequent variables as well as with a lag (Hamilton, 1994). The variables that enter later only affect the earlier variables with a lag. In other words, earlier series in the system are considered to be more exogenous than the subsequent ones. For the model in (5-7), the variable ordering may be chosen based on a priori knowledge on the structure of the relationship between the variables.

As mentioned above, the variables of interest (personal income growth, earning shares of resource and manufacturing sectors, and number of the big-box retailers) are thought to have an underlying endogenous relationship. As defined by the BEA, personal income is the income received by all persons from all sources. Therefore, personal income growth can be assumed to be the most exogenous and thus comes first in the variable ordering. The resource earnings come second in the ordering because it is composed of the county's earnings share in the primary sector of the economy. The manufacturing earnings share comes third since it belongs to the secondary sector of the economy. The number of big-box retailers, comprised of the sum of Wal-Mart and Target stores, comes last in the ordering as retailing fits into the tertiary sector of the economy, which means that the big-box retailer is assumed to be the most endogenous variable in the model. This variable ordering implies that the effect of the big-box retailers on other variables (i.e., personal income growth) in the model may take effect with at least one lag. In sum, the analysis in this chapter uses the following variable ordering:

(5-8) $$(gwth_{it}, res_{it}, mnf_{it}, bx_{it}),$$

where $gwth_{it}$ is the personal income growth in county $i$ at time $t$, $res_{it}$ is the share of resource earnings, $mnf_{it}$ is the share of manufacturing sector earnings, and $bx_{it}$ is the number of the big-box retailer stores.

## 5.4. Empirical Results

### *5.4.1. Panel Unit Root Test*

To use the panel VAR approach, all variables need to be stationary or integrated of the order one. Hence, testing the unit root is the first phase of the analysis. There are two classes of tests that can be used to detect the presence of the unit roots in the panel data. The first-generation panel unit root tests by Hadri (2000) has been developed assuming cross-section independence across units in the panel (with the exception of common time effects). In second-generation tests, the assumption of cross-sectional independence is relaxed, which allows for an array of dependence among the different units (Pesaran, 2007, Smith, et al., 2004).

The Fisher’s test as suggested in Maddala and Wu (1999) is used to test for the presence of unit root in the variables. The advantage of this test as Maddala and Wu (1999) point out is that it does not require a balanced panel as in the case of the IPS test. The test also allows for heterogeneity in the autoregressive coefficient of the Dickey-Fuller regression, which disregards cross-sectional dependence in the data.

Based on the p-values of individual unit root tests, the Fisher's test assumes that all series are non-stationary under the null hypothesis against the alternative that at least one series in the panel is stationary. The results for the unit root tests are shown in Table 5.2. The results suggest that all variables are stationary except for the number of the big-box stores, which is integrated of order one. Therefore, all other variables except for $bx_{it}$ enter the model in levels.

**Table 5.2. Panel Unit Root Tests**

| Variable | Augmented Dickey-Fuller | Phillips-Perron |
|---|---|---|
| *Income growth* | | |
| Level | 25400*** | 55800*** |
| Difference | 58700*** | 126000*** |
| *Share of resource earnings* | | |
| Level | 10300*** | 17100*** |
| Difference | 28400*** | 64900*** |
| *Share of manufacturing earnings* | | |
| Level | 8615*** | 11000*** |
| Difference | 23000*** | 50300*** |
| *Number of big-box stores* | | |
| Level | 3555 | 3607 |
| Difference | 8676*** | 17300*** |

*Notes:* All unit root tests are performed with 1 lag and a trend. (*), (**), (***) represents significance at the 10%, 5% and 1% level respectively.

### *5.4.2. Panel VAR Estimation Results*

To estimate the model in equation (5-7), a panel VAR package, PVAR, in Stata13 first developed in Love and Zicchino (2006) is utilized. The optimal lag length for the model is selected through moment model selection criteria (MMSC) developed by Andrews and Lu (2001). Table 5.3 reports the MMSC Bayesian information criterion (MBIC), MMSC Akaike's information criterion (MAIC), and MMSC Hannan and Quinn information criterion (MQIC).

Similar to maximum likelihood-based information criteria (i.e. AIC, BIC and HQIC), the model which minimizes the MAIC, MBIC or MQIC is the preferred model. Consequently, for the panel VAR mode the optimal lag length is 4. Note that for just identified systems like in (3-7) the Hansen's (1982) J statistic is equal to the MAIC, MBIC and MQIC.

**Table 5.3. Lag Order Selection Criteria**

| Lag | CD | J | J *p-value* | MBIC | MAIC | MQIC |
|---|---|---|---|---|---|---|
| 1 | 0.9897 | 1.75e-29 | 0.00 | 1.75e-29 | 1.75e-29 | 1.75e-29 |
| 2 | 0.9812 | 9.17e-29 | 0.00 | 9.17e-29 | 9.17e-29 | 9.17e-29 |
| 3 | 0.9921 | 7.43e-29 | 0.00 | 7.43e-29 | 7.43e-29 | 7.43e-29 |
| 4 | 0.9937 | 5.05e-30 | 0.00 | 5.05e-30* | 5.05e-30* | 5.05e-30* |
| *No. of obs.* | 51,799 | | | | | |
| *No. of panels* | 3,047 | | | | | |
| *Average no. of T* | 17 | | | *Sample:* | 1988 - 2004 | |

*Note:* These statistics are produced using the *pvarsoc*, a Stata module which reports the coefficient of determination (CD), J statistics as in Hansen (1982) and corresponding *p-value* (J *pvalue*). Also this table reports the moment model selection criteria developed by Andrews and Lu (2001): MMSC-Bayesian information criterion (MBIC), MMSC-Akaike's information criterion (MAIC), and MMSC-Hannan and Quinn information criterion (MQIC).

The panel VAR estimation results are presented in Table 5.4. The results show that the big-box stores have a negative impact on personal income growth. This is consistent with the results obtained in the previous chapter and the negative impact of the big-box stores expansion on retail wage as discussed in chapter 3. The estimated coefficients are significant for the first two lags. Jantzen, Pescatrice and Braunstein (2009) analogously describes a negative effect on personal income as a result of Wal-Mart's growth (i.e., sales increases).

On the other hand, personal income growth has a positive impact on big-box store location decisions. This is consistent with the claims that these big-box retailers tend to locate in areas exhibiting a higher income growth. As per income convergence theory, this would imply that these big-box stores tend to locate in poorer regions (higher growth counties). Nonetheless, the coefficients are not statistically significant and turn negative after the third lag.

**Table 5.4. Panel VAR Estimates**

| Response of | | Response to | | | |
|---|---|---|---|---|---|
| | | $growth_{it}$ | $res_{it}$ | $mnf_{it}$ | $\Delta bx_{it}$ |
| $growth_{it}$ | | | | | |
| | L1. | -0.3249 | -0.0524 | 0.0783 | -0.0010 |
| | | [-25.78]*** | [-3.89]*** | [15.80]*** | [-2.19]** |
| | L2. | 0.294 | -0.0152 | 0.0179 | -0.0009 |
| | | [2.33]** | [-1.11] | [3.78]*** | [-1.89]* |
| | L3 | 0.0210 | 0.0026 | -0.0027 | -0.0004 |
| | | [1.78]* | [0.22] | [-0.58] | [-1.15] |
| | L4 | 0.0013 | -0.0087 | 0.0414 | 0.0000 |
| | | [0.10] | [-0.84] | [11.29] | [-0.07] |
| $res_{it}$ | | | | | |
| | L1. | 0.0013 | 0.5500 | -0.0056 | 0.010 |
| | | [0.33] | [13.50]*** | [-2.64]** | [2.27]** |
| | L2. | -0.0029 | 0.1291 | 0.0011 | 0.0007 |
| | | [-0.57] | [3.23]*** | [0.54] | [2.08]** |
| | L3 | 0.0032 | 0.0127 | -0.0002 | 0.0001 |
| | | [0.78] | [0.37] | [-0.08] | [0.15] |
| | L4 | -0.0010 | 0.0567 | 0.0013 | 0.0010 |
| | | [-0.24] | [2.14]** | [0.64] | [2.41]** |
| $mnf_{it}$ | | | | | |
| | L1. | 0.0136 | 0.0019 | 0.6888 | 0.0013 |
| | | [2.28]** | [0.21] | [44.47]*** | [2.83]*** |
| | L2. | 0.0217 | 0.0056 | 0.1065 | 0.0001 |
| | | [3.11]*** | [0.58] | [6.31]*** | [0.12] |
| | L3 | 0.0454 | 0.0113 | 0.0517 | 0.0016 |
| | | [6.80]*** | [1.37] | [3.59]*** | [2.53]** |
| | L4 | 0.0126 | 0.0083 | 0.0120 | 0.0003 |
| | | [2.33]** | [1.28] | [1.15] | [0.56] |
| $\Delta bx_{it}$ | | | | | |
| | L1. | 0.0295 | -0.1269 | -0.0089 | -0.0057 |
| | | [0.54] | [-1.32] | [-0.18]** | [-0.76] |
| | L2. | 0.0183 | 0.1059 | 0.0315 | -0.0132 |
| | | [0.27] | [0.93] | [0.68] | [-1.92]* |
| | L3 | 0.0190 | -0.1454 | -0.0137 | -0.0035 |
| | | [0.33] | [-1.99]** | [-0.37] | [-0.43] |
| | L4 | -0.0556 | 0.0185 | 0.0329 | -0.0103 |
| | | [-0.91] | [0.30] | [0.86] | [2.15]** |
| *No. of obs.* | 42658 | | | | |
| *No. of panels* | 3047 | | | | |

*Notes:* Variable definitions are in Table 5.1. The four-variable VAR model is estimated by system GMM, fixed effects are removed prior to estimation (see Methodology section for more details). Reported numbers show the coefficients of regressing the row variables on 4 lags of the column variables. Lag selection criteria follows the model selection criteria in Table 5.3. T-statistics are in brackets. ***, ** and * indicates significance at 1%, 5% and 10%, respectively. Notice the Wal-Mart and Target variables enter the model in first difference as they are both integrated of order one.

The results in Table 5.4 also show how the big-box stores have a significant positive impact on both resource and manufacturing earnings share. Moreover, the share of resource earnings has a negative and statistically significant impact on income growth. This is consistent with the natural resource curse, i.e. natural resource dependence tends to be associated with lower economic growth, as discussed in James and Aadland (2011). But this doesn't support the findings in Chapter 4 of this dissertation. In contrast, a larger share of manufacturing earnings has a significant and positive effect on personal income growth.

### *5.4.3. Impulse Response Functions*

To better evaluate the dynamics of the effects estimated in Table 5.4, impulse response functions (IRF) are produced (Figure 5.1). The impulse response magnitudes are presented in Table 5.5. The orthogonalized IRFs using the variable sequence ordering in (5-8) show the response of one variable of interest (i.e., personal income growth) to an orthogonal shock in another variable of interest (i.e., the big-box stores). The orthogonalization of the responses enables the identification of the impact of one shock at a time while holding other shocks constant (Hamilton, 1994).

The model in (5-7) is a four-variable VAR model with four-equations. As a result, there are sixteen IRFs for the system. A shock equal to one standard deviation of its residual is applied to each variable in the system while holding all other variables' innovation constant. The IRF graphs display how each variable responds to each of those shocks. The horizontal axis shows the time period elapsed (i.e., years) after the shock is applied. The y-axis indicates the size and direction of the impulse. The $\pm 2$ standard error confidence bounds for variables' responses are constructed through the Monte Carlo simulation and are denoted by the short-dashes.

As it can be seen in Figure 5.1, in the top right corner, personal income growth responds negatively to a one standard deviation shock to the change in the number of big-box stores.

Although quite small, the effect peaks around the first period and quickly vanishes afterwards. On the flip side, a shock to personal income growth has an insignificant effect on the change in the number of big box stores (as seen in the bottom left corner in Figure 5.1).

A shock to the county share of resource earnings has a negative and significant effect on income growth. As discussed above, this is consistent with the natural resources curse argued in James and Aadland (2011) but not consistent with Chapter 4 in the dissertation. The effect while statistically significant, is very small in terms of magnitude (Table 5.5) and gradually dissipates after the first period.

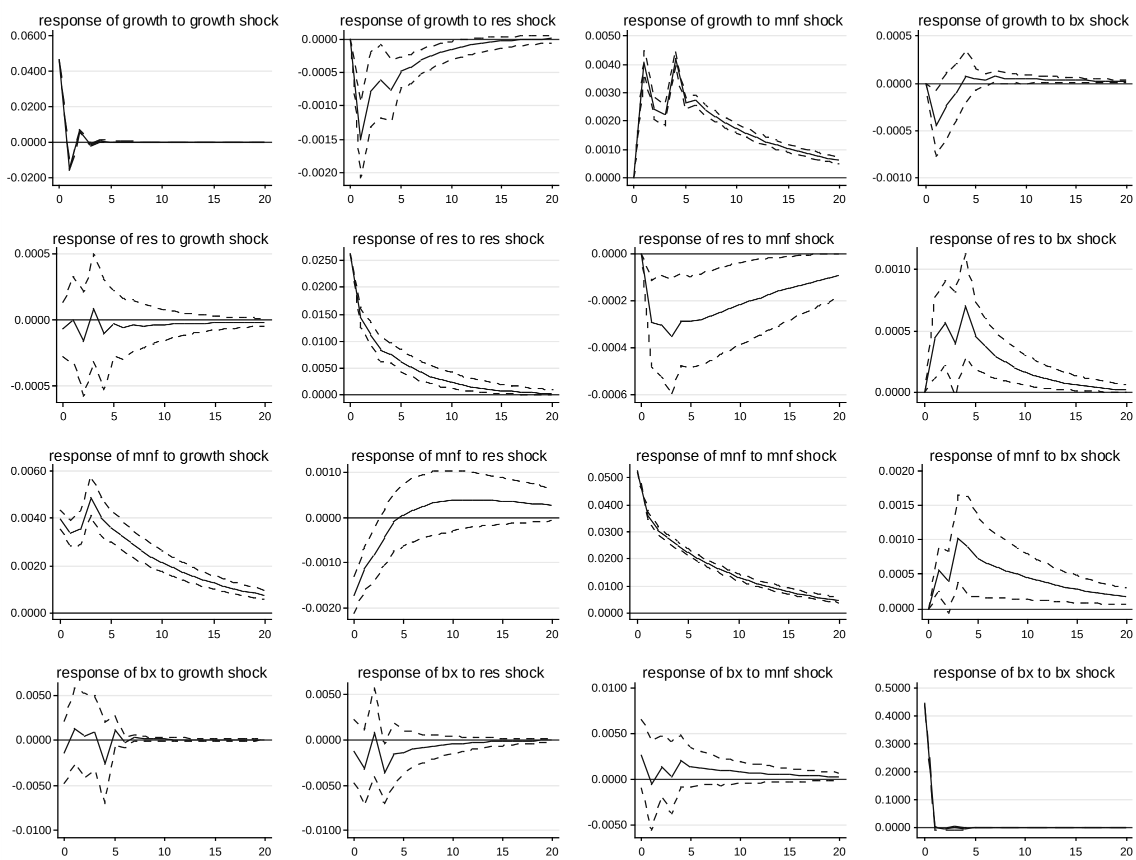


**Figure 5.1. Impulse response functions**

Note: Every row presents the different shocks to personal income growth (growth), resource earnings share (res), manufacturing earnings share (mnf), and the change in the number of big box stores (Δbx), respectively. Errors are 5% on each side generated by Monte-Carlo with 1000 repetitions.

**Table 5.5. Impulse Response Functions Magnitudes**

| | Time | *growth* | *resource* | *mnf* | *Δbx* |
|---|---|---|---|---|---|
| *growth* | 0 | 0.0467 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | -0.0149 | -0.0015 | 0.0041 | -0.0005 |
| | 2 | 0.0065 | -0.0008 | 0.0024 | -0.0002 |
| | 3 | -0.0012 | -0.0006 | 0.0022 | -0.0001 |
| | 4 | 0.0009 | -0.0008 | 0.0042 | 0.0001 |
| | 8 | 0.0003 | -0.0003 | 0.0022 | 0.0000 |
| | 12 | 0.0002 | -0.0001 | 0.0014 | 0.0000 |
| *res* | 0 | -0.0001 | 0.0261 | 0.0000 | 0.0000 |
| | 1 | 0.0000 | 0.0144 | -0.0003 | 0.0004 |
| | 2 | -0.0002 | 0.0113 | -0.0003 | 0.0006 |
| | 3 | 0.0001 | 0.0084 | -0.0004 | 0.0004 |
| | 4 | -0.0001 | 0.0077 | -0.0003 | 0.0007 |
| | 8 | 0.0000 | 0.0035 | -0.0002 | 0.0002 |
| | 12 | 0.0000 | 0.0016 | -0.0002 | 0.0001 |
| *mnf* | 0 | 0.0039 | -0.0017 | 0.0522 | 0.0000 |
| | 1 | 0.0034 | -0.0011 | 0.0360 | 0.0006 |
| | 2 | 0.0035 | -0.0008 | 0.0304 | 0.0004 |
| | 3 | 0.0049 | -0.0004 | 0.0276 | 0.0010 |
| | 4 | 0.004 | -0.0001 | 0.0250 | 0.0009 |
| | 8 | 0.0026 | 0.0003 | 0.0164 | 0.0006 |
| | 12 | 0.0017 | 0.0004 | 0.0108 | 0.0004 |
| *Δbx* | 0 | -0.0014 | -0.0013 | 0.0026 | 0.4426 |
| | 1 | 0.0014 | -0.0033 | -0.0005 | -0.0025 |
| | 2 | 0.0005 | 0.0009 | 0.0014 | -0.0059 |
| | 3 | 0.0008 | -0.0037 | 0.0003 | -0.0015 |
| | 4 | -0.0027 | -0.0016 | 0.0021 | -0.0046 |
| | 8 | 0.0002 | -0.0007 | 0.0009 | 0.0000 |
| | 12 | 0.0001 | -0.0003 | 0.0006 | 0.0000 |

*Note:* All variables are included in levels except for big-box that is included in differences. Each cell shows a response of the row variable to a shock in column variable (at a given time).

On the other hand, applying a shock to the share of earnings in manufacturing has a positive and highly significant impact on personal income growth. This implies that a higher share of manufacturing earnings is tied to a higher rate of growth for personal income. The effect peaks twice around the first and fourth period. It then gradually fades out over the rest of the time horizon. Moreover, the IRFs indicate that the big-box stores have a positive effect on the county share of resource earnings. Additionally, contrary to popular belief, the IRF shows that the big-box stores have a positive impact on the county share of manufacturing earnings. Nonetheless, the shocks to the share of resource and manufacturing earnings have no significant impact on big-box store location decision.

### *5.4.4. Variance Decompositions*

The variance decompositions are presented in Table 5.6. This table shows the percent variation in the row variable explained by the column variable. Note, only the total effect accumulated over 10 years is reported, but longer time horizons produced equivalent results. As shown in the table, the share of manufacturing earnings explains the highest variation in personal income growth, followed by resources and the big-box retailers (2.92%, 0.18%, and 0.011%, respectively). For the share of resource earnings, the presence of big-box stores explains the highest percent of total variation, followed by the share of manufacturing earnings, and income growth (0.14%, 0.06%, and 0.005%, respectively).

The highest percent of variation in the share of manufacturing earnings is explained by income growth (1.51%), followed by resources earnings (0.067%) and big-box retailers (0.052%). The percent of variation in the big-box location decision is better explained by the share of resource earnings (0.017%), followed by the share of earnings in manufacturing (0.01%) and income growth (0.007%). However, recall that according to panel VAR estimates and the IRFs none of these have a significant impact on the location decision.

**Table 5.6. Variance Decompositions**

| | growth | res | mnf | $\Delta$bx |
|---|---|---|---|---|
| growth | 0.96896 | 0.00175 | 0.02917 | 0.00011 |
| res | 0.00005 | 0.99802 | 0.00058 | 0.00136 |
| mnf | 0.01512 | 0.00067 | 0.98370 | 0.00052 |
| $\Delta$bx | 0.00007 | 0.00017 | 0.00010 | 0.99966 |

*Note:* Variation in the row variable explained by column variable (10 periods ahead).

## 5.5. Summary and Concluding Remarks

This chapter complements the analyses in the fourth chapter and earlier works in the literature by examining the dynamics effects from the presence of the big-box stores on personal income growth at the county level, over the period of 1987-2005. The county share of natural resources and manufacturing earnings are also included in the analysis. The analysis is done applying the panel vector autoregressive approach whereby the dynamic relationship among variables of interest is examined, assuming that all variables in the system are endogenous to one another. The calculation and analysis of impulse response functions (IRFs) allows us to separate the responses from personal income growth and big-box retailers to shocks from each of the variable of interest (orthogonalized impulse-response functions). The orthogonalization of the responses enables the identification of the impact of one shock at a time while holding other shocks constant. Variance decompositions are also reported, which show the percent variation in one variable that is explained by the shock or innovation to another variable accumulated over time.

One of the key assumptions in the third essay is the variable ordering for the panel VAR. The personal income growth is assumed to be the most exogenous (comes first in the variable ordering) and the big-box retailers are most endogenous (come last in the variable ordering). The results suggest that the big-box retailers negatively affect personal income growth while the impact of the personal income growth on the big-box retailers is insignificant. Furthermore, the results indicate that the big-box retail stores have a positive effect on the share of resource earnings. In addition, contrary to the popular belief, the big-box retailers have a positive impact on the share of manufacturing earnings, too. However, the shares of resource and manufacturing earnings have no significant impact on the big-box stores' location decision.

# CHAPTER 6

# SUMMARY AND CONCLUSIONS

Throughout the years, big-box retail stores such as Wal-Mart and Target have become the focus of many studies researching their impact on local economies - including retail wages, retail employment and economic growth (personal income growth). This dissertation attempts to examine three closely related topics: big-box retailers, regional economy and personal income growth.

Specifically, this dissertation studies (i) the dynamic interrelationship among the presence of big-box retail stores, retail wage and retail employment, (ii) the impact of big-box retailers on personal income growth, and (iii) the dynamic interrelationship between the presence of big-box retailers and personal income growth. The research draws important insights with potential policy implications for regional developers and policy makers. To achieve the research goals, the numbers of Wal-Mart and Target stores across U.S. counties within the 48 contiguous states are compiled. Also, personal income, socio-demographic data and earning shares of industry are compiled for the empirical analysis.

The first essay attempts to investigate the dynamic interactions generated by the presence of the big-box retailers, with respect to retail wages, and retail employment at the county level. This is studied for the period of 1986-2005 using the vector autoregressions on panel data (panel VAR), impulse response functions (IRFs) and variance decompositions. The panel VAR model is useful when endogeneity and unobserved heterogeneity are present, and it provides a unifying empirical framework and identification strategy that has been absent in the previous literature.

The empirical results in the first essay show that the presence of the big-box retailers increases the number of retail jobs[13] in a county while it decreases the level of retail wages. The effect on retail employment is relatively larger than the effect on retail wage in terms of the total variation explained by the big-box retailers. These effects are mostly driven by Wal-Mart. Target's effect on retail wage and employment are (statistically) insignificant.[14]

The big-box retailers' location decision is also of interest. The variance decompositions show that Target's location decision is slightly more heavily affected by retail wage than employment level. Conversely, retail employment has a relatively bigger impact on Wal-Mart's location decision compared to retail wages. As expected, Target is characterized by a follower strategy regarding its location decision, meaning that Target may open a store in a county where a Wal-Mart store exists. On the contrary, Wal-Mart tries to avoid counties with an existing Target store.

The second essay attempts to investigate the impact of big-box retailers on personal income growth, using the neoclassical growth models of cross-country income convergence. Whether the big-box retailers' presence, particularly Wal-Mart and Target, have a negative impact on local economic growth has been a permeating question amongst regional developers, policy makers and researchers. The analysis is performed using cross-sectional county level data between 2000 and 2005 for the 48 contiguous states.

The results indicate that counties with both Wal-Mart and Target stores have experienced slower growth in personal income between 2000 and 2005, even after controlling for spatial dependencies and for the dot-com bubble in early 2000s. The presence of Wal-Mart seems to drive

[13] The exact source for these gains in employment in the retail sector is unknown. However, some complementary establishments may yield positive benefits from Wal-Mart's presence , which may have an impact on retail job gains.
[14] Note that firms with labor unions generally have no impact on wages. This may be the reason why Target's effect on retail wage is statistically insignificant.

the impact. The impact of Target is also negative but it becomes statistically insignificant when the spatial autocorrelation is controlled for. Even though the spatial model improves the fit of the (theoretical) model, further diagnostics on the spatial model specification indicate some possible remaining misspecifications issues. Presumably, possible endogeneity existing between the big-box retailers' location decision and personal income growth may be a source of misspecification.

The third essay complements the analysis in the second essay. In the third essay, the dynamic interrelationship among the big-box retailers and personal income growth is examined using county level data for the period of 1987-2005 and the panel VAR. The county's share of natural resource and manufacturing earnings are also included in the analysis. One of the key assumptions in the third essay is the variable ordering for the panel VAR. The personal income growth is assumed to be the most exogenous (comes first in the variable ordering) and the big-box retailers are the most endogenous (come last in the variable ordering).

The findings suggest that the big-box retailers are negatively related to the personal income growth while the impact of the personal income growth on the big-box retailers is (statistically) insignificant. Furthermore, the results also indicate that these big-box stores have a positive effect on the share of resource earnings. In addition, contrary to the popular belief, the big-box retailers have a positive impact on the share of manufacturing earnings, too. However, the shares of resource and manufacturing earnings have no significant impact on the big-box stores' location decision.

In sum, using advanced econometric approaches such as the panel VAR and spatial econometrics, this dissertation arrives at the following conclusion: (i) the presence of the big-box retailers, increases retail jobs while it decreases retail wages. The effect on retail employment is slightly larger. Wal-Mart drives these effects while the presence of Target is insignificant, (ii) Counties with big-box retailers have experienced slower growth in personal income between 2000 and 2005 - even after controlling for spatial autocorrelation and for the dot-com bubble in early 2000s, and the impact of Wal-Mart's presence is stronger, and (iii), these big-box retailers also have a negative impact on the personal income growth over time whereas personal income growth has an inconsequential effect on the number of the big-box retailers in a region.

Based on the findings of this research, the presence of the big-box retailers has both negative and positive impacts on the local economy, i.e., more jobs but lower wages (in the retail sector), increasing share of earnings in resource and manufacturing sectors, but decreasing income growth. Policy implications of this study are obvious, minimize the negative impacts and maximize the positive impacts. Regional developers and policy makers should focus on ensuring that competitive wages are offered by big-box chain retailers entering their regions to ensure competition. Not surprisingly, according to recent news and a tightening labor market in the retail sector, big box stores, i.e., Wal-Mart, Target, and TJ-Maxx, another big-box retailer, are already ceding to the pressure to increase wages (Benn Steil, 2015, Isidore, 2015, Lynch, 2015, Ramakrishnan, 2015, Stangler, 2015). As implied in these publications, public relations pressures, political interest and legal necessities may better explain the pay hikes. Future research should focus on the economic effects of such raises in wages in the local retail market and personal income growth.

In interpreting the empirical results of the big-box retailers' effect, a couple of caveats are to be noted. First, the impact of the big-box retailers in the IRFs should not be interpreted as the magnitude of changes. The IRFs depict graphically and offer a visual impression of the dynamic interrelationships within the system. Second, the behavior of the residuals from the panel VAR are not further tested to ensure they follow the assumptions, and thus the confidence bands from the Monte Carlo simulation may not be accurate. When the residuals from the panel VAR have serial correlation and heteroscedasticity problems, the confidence bands in the IRFs may be biased. This implies that the insignificant impact of Target on regional economy might not be accurate.

# BIBLIOGRAPHY

Acemoglu, D., S. Johnson, and J.A. Robinson. 2003. *An African Success Story: Botswana*. Princeton: Princeton University Press.

Ailawadi, K.L., J. Zhang, A. Krishna, and M.W. Kruger. 2010. When Wal-Mart enters: How incumbent retailers react and how this affects their sales outcomes. *Journal of Marketing Research* 47:577-593.

Andrews, D.W., and B. Lu. 2001. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics* 101:123-164.

Anselin, L. 2004. Exploring spatial data with GeoDaTM: a workbook. *Urbana* 51:61801.

---. 2001. Rao's score test in spatial econometrics. *Journal of statistical planning and inference* 97:113-139.

---. 1988. *Spatial econometrics: Methods and models*: Studies in Operational Regional Science series Norwell, Mass.; London and Dordrecht: Kluwer Academic.

Anselin, L., and A.K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. *statistics textbooks and monographs* 155:237-290.

Arellano, M., and O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68:29-51.

Artz, G.M., and K.E. Stone. 2006. Analyzing the impact of Wal-Mart supercenters on local food store sales. *American journal of agricultural economics* 88:1296-1303.

Barnes, N.G., A. Connell, L. Hermenegildo, and L. Mattson. 1996. Regional differences in the economic impact of Wal-Mart. *Business Horizons* 39:21-25.

Barro, R.J., and X. Sala-i-Martin. 1992. Convergence. *Journal of Political Economy* 100:223-251.

Basker, E. 2007. The causes and consequences of Wal-Mart's growth. *The Journal of Economic Perspectives* 21:177-198.

---. 2011. Does Wal-Mart Sell Inferior Goods? *Economic Inquiry* 49:973-981.

---. 2005. Job creation or destruction? Labor market effects of Wal-Mart expansion. *Review of Economics and Statistics* 87:174-183.

---. 2005. Selling a cheaper mousetrap: Wal-Mart's effect on retail prices. *Journal of Urban Economics* 58:203-229.

---. 2007. When good instruments go bad: A reply to Neumark, Zhang, and Ciccarella. *Zhang, and Ciccarella (April 2007)*.

Basker, E., S. Klimek, and P. Hoang Van. 2012. Supersize It: The Growth of Retail Chains and the Rise of the "Big-Box" Store. *Journal of Economics & Management Strategy* 21:541-582.

Benn Steil, D.W. 2015. Why Did Walmart Raise Its Wages. Forbes. http://www.forbes.com/sites/realspin/2015/04/02/why-did-walmart-raises-its-wages/.

Bernstein, J., and L.J. Bivens. 2006. The Wal-Mart debate: A false choice between prices and wages. *EPI Issue Brief* 223:15.

Bloom, D.E., D. Canning, and P.N. Malaney. 2000. Population dynamics and economic growth in Asia. *Population and Development Review* 26:257-290.

Bonanno, A., and S.J. Goetz. 2012. WalMart and Local Economic Development: A Survey. *Economic Development Quarterly* 26:285-297.

Bonanno, A., and R.A. Lopez. 2012. Wal-Mart's monopsony power in metro and non-metro labor markets. *Regional Science and Urban Economics* 42:569-579.

Carlino, G.A., and E.S. Mills. 1987. The Determinants of County Growth. *Journal of Regional Science* 27:39-54.

Cliff, A., and K. Ord. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4:267-284.

Deller, S.C., T.-H.S. Tsai, D.W. Marcouiller, and D.B. English. 2001. The role of amenities and quality of life in rural economic growth. *American journal of agricultural economics* 83:352-365.

Dixit, A. 1973. The optimum factory town. *The Bell Journal of Economics and Management Science*:637-651.

Drewianka, S., and D. Johnson. 2006. Wal-Mart and Local Labor Markets, 1990-2004. Working Paper, University of Wisconson.

Dube, A., T.W. Lester, and B. Eidlin. 2007. Firm entry and wages: impact of Wal-Mart growth on earnings throughout the retail sector. *Available at SSRN 841684*.

Fisher, R.A. 1932. *Statistical methods for research workers*. 4th ed. Edinburgh: Oliver & Boyd.

Foster, L., J. Haltiwanger, and C.J. Krizan. 2006. Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s. *The Review of Economics and Statistics* 88:748-758.

Freeman, R. 2003 Wal-Mart collapses US cities and towns. *Executive Intelligence Review*, November 21, pp. 7-9.

Goetz, S.J., and S.S. Shrestha. 2009. Explaining Self-Employment Success and Failure: Wal-Mart Versus Starbucks, or Schumpeter Versus Putnam. *Social Science Quarterly* 90:22-38.

Goetz, S.J., and H. Swaminathan. 2006. Wal-Mart and County-Wide Poverty. *Social Science Quarterly* 87:211-226.

Hadri, K. 2000. Testing for stationarity in heterogeneous panel data. *The Econometrics Journal* 3:148-161.

Haltiwanger, J., R. Jarmin, and C.J. Krizan. 2010. Mom-and-pop meet big-box: Complements or substitutes? *Journal of Urban Economics* 67:116-134.

Hamilton, J.D. 1994. *Time series analysis*. 1 ed. Princeton: Princeton University Press.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 50:1029-1054.

Hausman, J., and E. Leibtag. 2007. Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart. *Journal of Applied Econometrics* 22:1157-1177.

Hicks, M.J. 2008. Estimating Wal-Mart's impacts in Maryland: A test of identification strategies and endogeneity tests. *Eastern Economic Journal* 34:56-73.

---. 2007. Job turnover and wages in the retail sector: The influence of Wal-Mart. *Journal of Private Enterprise* 22:137-160.

Hicks, M.J., and K.L. Wilburn. 2001. The regional impact of Wal-Mart entrance: A panel study of the retail trade sector in West Virginia. *The Review of Regional Studies* 31:305-313.

Higgins, M.J., D. Levy, and A.T. Young. 2006. Growth and convergence across the United States: evidence from county-level data. *The Review of Economics and Statistics* 88:671-681.

Holmes, T.J. 2011. The Diffusion of Wal-Mart and Economies of Density. *Econometrica* 79:253-302.

Hurlin, C., and B. Venet. 2003. Granger causality tests in panel data models with fixed coefficients. Unpublished manuscript, University of Paris IX.

Irwin, E.G., and J. Clark. 2006. Wall Street vs. Main Street: What are the Benefits and Costs of Wal-Mart to Local Communities? Choices Magazine. http://www.choicesmagazine.org/2006-2/grabbag/2006-2-14.htm.

Isidore, C. 2015. Walmart ups pay well above minimum wage. CNN Money. CNN. http://money.cnn.com/2015/02/19/news/companies/walmart-wages/.

James, A., and D. Aadland. 2011. The curse of natural resources: An empirical investigation of US counties. *Resource and Energy Economics* 33:440-453.

Jantzen, R., D. Pescatrice, and A. Braunstein. 2009. Wal-Mart and the US economy. *Eastern Economic Journal* 35:297-308.

Jia, P. 2008. What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. *Econometrica* 76:1263-1316.

Keil, S.R., and L.C. Spector. 2005. The Impact of Wal-Mart on Income and Unemployment Differentials in Alabama. *Review of Regional Studies* 35:336-355.

Ketchum, B.A., and J.W. Hughes. 1997. Wal-Mart and Maine: The effect on employment and wages. *Maine Business Indicators* 42:6-8.

Kolko, J., and D. Neumark. 2010. Does local business ownership insulate cities from economic shocks? *Journal of Urban Economics* 67:103-115.

Koutsomanoli-Filippaki, A., and E. Mamatzakis. 2009. Performance and Merton-type default risk of listed banks in the EU: A panel VAR approach. *Journal of Banking & Finance* 33:2050-2061.

LeSage, J., and R.K. Pace. 2009. *Introduction to spatial econometrics*: CRC press.

LeSage, J.P. 2014. What Regional Scientists Need to Know about Spatial Econometrics. *Review of Regional Studies* 44:13-32.

Love, I., and R.T. Ariss. 2014. Macro-financial linkages in Egypt: A panel analysis of economic shocks and loan portfolio quality. *Journal of International Financial Markets, Institutions and Money* 28:158-181.

Love, I., and L. Zicchino. 2006. Financial development and dynamic investment behavior: Evidence from panel VAR. *The Quarterly Review of Economics and Finance* 46:190-210.

Lucas, R.E., Jr. 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics* 22:3-42.

Lynch, D. 2015. Target Minimum Wage Raise 2015: All Employees To Be Paid $9 Per Hour Beginning April. International Business Times. http://www.ibtimes.com/target-minimum-wage-raise-2015-all-employees-be-paid-9-hour-beginning-april-1851848?rel=rel2.

Maddala, G.S., and S. Wu. 1999. A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics* 61:631-652.

Malmberg, B. 1994. Age structure effects on economic growth - Swedish evidence. *Scandinavian Economic History Review* 42:279-295.

Mankiw, N.G., D. Romer, and D.N. Weil. 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107:407-437.

McGee, J.E., and G.G. Gresham. 1996. When Wal-Mart comes to town: A look at how local merchants respond to the retailing giant's arrival. *Journal of Business and Entrepreneurship* 8:43-52.

Neumark, D., J. Zhang, and S. Ciccarella. 2008. The effects of Wal-Mart on local labor markets. *Journal of Urban Economics* 63:405-430.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*:1417-1426.

Ostrander, A.P. 2011. The Expansion of a Retail Chain: An Analysis of Wal-Mart Locations in the United States. MS Thesis, University of North Texas.

Ozment, J., and G. Martin. 1990. Changes in the competitive environments of rural trade areas: Effects of discount retail chains. *Journal of Business Research* 21:277-287.

Pesaran, M.H. 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22:265-312.

Quigley, J.M. 1998. Urban diversity and economic growth. *The Journal of Economic Perspectives*:127-138.

Quinn, B. 2005. *How Walmart Is Destroying America And The World And What You Can Do About It*: Random House LLC.

Ramakrishnan, S. 2015. Wal-Mart to Raise Wages for 100,000 U.S. Workers in Some Departments. Thomson Reuters. Reuters. http://www.reuters.com/article/2015/06/02/us-wal-mart-stores-wages-idUSKBN0OI1EW20150602.

Rupasingha, A., S.J. Goetz, and D. Freshwater. 2002. Social and institutional factors as determinants of economic growth: Evidence from the United States counties. *Papers in regional Science* 81:139-155.

Schuetz, J. 2015. Why Are Wal-Mart and Target Next-Door Neighbors? *Regional Science and Urban Economics* 54:38-48.

Semykina, A., and J.M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157:375-380.

Slater, R. 2003. *The Wal-Mart Decade: How a New Generation of Leaders Turned Sam Walton's Legacy Into the World's # 1 Company*: Portfolio

Smith, L.V., S. Leybourne, T.H. Kim, and P. Newbold. 2004. More powerful panel data unit root tests with an application to mean reversion in real exchange rates. *Journal of Applied Econometrics* 19:147-170.

Stangler, C. 2015. Wal-Mart, TJX, Target Raise Wages: What's Behind The Pay Hikes In Retail. International Business Times. http://www.ibtimes.com/wal-mart-tjx-target-raise-wages-whats-behind-pay-hikes-retail-1856022.

Stone, K.E. 1988. The effect of Wal-Mart Stores on businesses in host towns and surrounding towns in Iowa. Working Paper, Iowa State University.

---. 1997. Impact of the Wal-Mart phenomenon on rural communities. *Increasing understanding of public problems and policies* 1997:1-21.

---. 1995. Impact of Wal-Mart stores on Iowa communities: 1983-93. *Economic Development Review* 13:1-60.

Target. 2015. Corporate Fact Sheet. Target Press Room. http://pressroom.target.com/corporate.

Wal-Mart Stores, I. 2015. Walmart Corporate Website: Our Story. http://corporate.walmart.com/our-story/locations#/united-states.

Wooldridge, J. 2015. *Introductory Econometrics: A Modern Approach*. 6 ed: South-Western College Pub.

Zak, P.J., and S. Knack. 2001. Trust and Growth. *The Economic Journal* 111:295-321.

# CURRICULUM VITAE

Denis Jose Peralta
(March 2016)

EDUCATION

PhD in Economics, Applied Economics, Utah State University, Logan Utah. (expected 5/16).

MS in Applied Economics, Utah State University, Logan Utah. (12/10).

Master of Business Administration (MBA), Utah State University, Logan Utah. (5/10).

BS in Industrial Engineering, Universidad Tecnologica de Santiago, Santo Domingo, D.R. (10/07).

EXPERIENCE

*Research Analysist*, Applied Economics, Utah State University, Logan, UT. Sept. 2011 – Jan. 2016

- Performed various econometric and statistical modelling and data analyses using STATA, R, MATLAB, Excel, SAS
- Identified, compiled, and organized large databases and participated in economic research discussions, proposal writing and literature reviews.
- Supported a team of researchers on modeling and data analysis using Stata, SPSS and Excel macros to clean and map data for further modeling and estimation
- Assisted on improving household survey research design and writing literature reviews.

*Instructor of Environmental Economics*, Applied Economics, Utah State University, Logan, UT.

- Taught class content and supported learning of students for master level Environmental Economics

*Business Customer Support Specialist*, Convergys, Logan, UT. April 2011 – August 2011

- Provided excellent customer and technical support to Comcast business clients

*Business Conference Coordinator*, Partners in Business, Utah State University, Logan UT. April 2010 – April 2011

- Maintained database and mapping of partners, speakers and sponsors
- Led and coordinated the Operational Excellence and Information Technology conferences
- Participated in sales, outreach, marketing, logistics and training of conferences coordinators

*Technical Support Specialist*, US Robotics, Santo Domingo, Dominican Republic. May 2007 – Sept. 2007

- Provided excellent customer service and technical support to U.S. market

*Quality Control Intern*, Cartones del Caribe, Santo Domingo, Dominican Republic. May 2006 – Aug. 2006

- Applied lean and statistical process control tools and techniques to reduce variation in various production processes and developed and updated standard operating procedures

RESEARCH

Peralta, D.J. (2016) *Three Essays on Big-Box Retailers and Regional Economics*, Unpublished Dissertation, Utah State University, Logan, UT.

Kim, M-K, D.J. Peralta, and B.A. McCarl (2014) Land-based greenhouse gas emission offset and leakage discounting. *Ecological Economics* 105: 265-273.

HONORS AND AWARDS

- The Honor Society of Phi Kappa Phi (inducted April 2010)
- Beta Gamma Sigma Business Honor Society (inducted April 2010)
- Member of Golden Key International Honor Society (inducted Sept. 2009)